DNA methylation in metatherian mammals roles in X chromosome inactivation & early development

Leeke, Bryony

Awarding institution:
King's College London

DNA methylation in metatherian mammals: roles in X chromosome inactivation & early development

Bryony Jane Leeke

King's College London
and
The Francis Crick Institute
PhD Supervisor: Dr James Turner

A thesis submitted for the degree of
Doctor of Philosophy
King's College London
May 2020

Declaration

I, Bryony Jane Leeke confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

At fertilisation, terminally differentiated gametes combine to form a totipotent zygote, from which all subsequent cell types of the developing embryo are derived. Control of gene expression via epigenetic regulation represents an important mechanism contributing to the attainment of totipotency, and subsequent specification of cell lineages. During embryonic development, there are dynamic changes to the epigenetic systems controlling gene expression.

Mammalian embryos differ from the embryos of other vertebrates in specifying an additional cell lineage that will lead to the development of a placenta. In addition, the evolution of mammalian XY sex chromosomes has resulted in the special requirement for dosage compensation of X-linked genes. The evolution of both the XY sex chromosomes and the placenta occurred prior to the divergence of the metatherian and eutherian lineages. Metatherian mammals are therefore perfectly placed in this evolutionary lineage to answer key questions about the evolution of mammalian traits.

In this thesis, I have examined the role of one key component of epigenetic regulation, DNA methylation, in the early embryos and adults of the metatherian *Monodelphis domestica*, with the aim of understanding how the roles for DNA methylation in X-chromosome inactivation and early development have evolved in the mammalian lineage.

I profiled the DNA methylation landscape of the *M. domestica* genome, and revealed that the inactive X chromosome displays a unique hypomethylated state. Using allele-specific analyses, I identified several novel X chromosome inactivation (XCI) escape genes, and demonstrated that XCI-escape genes are discernible by a unique DNA methylation signature. Finally, I profiled DNA methylation of *M. domestica* gametes and preimplantation embryos. I show DNA methylation at the region around *RSX,* the long non-coding RNA implicated in controlling metatherian XCI, to be a putative imprint controlling paternal XCI. Lastly, I discovered that unlike eutherian mammals, metatherian embryos do not undergo a dramatic genome-wide reprogramming of DNA methylation.

Acknowledgement

Many people have provided support and encouragement during this PhD; that my thanks extend to so many demonstrates my fortune in having such wonderful friends, colleagues, and family.

Firstly, to James. When I began to look for a PhD position, I knew that I wanted to work for someone who would always be in my corner. In joining your group, that is exactly what I found. The lab environment you have created fosters scientific creativity, intellectual rigor, and quite a bit of fun. Thank you for your mentorship, and for entrusting this project to me.

Thanks also to the entire Turner lab for support, discussion, and friendship. In particular, Fanny, who started this project with me and was my first guide. Jasmin, thank you for being an enthusiastic bioinformatics instructor and supportive friend. Daniel, thanks for excellent long-winded discussions of both minutiae and grand ideas. Sugako and Shantha both gave me generous help in many ways. Everyone else in the Turner lab, thank you for making it so fun to go to work; Taka, Val, Elias, Sergio, Andrew, Mahesh, Oana, Ruta, Jeremie and Haskan. And Charlotte, with whom I've shared this whole journey. Thank you for always proof-reading my emails, and for teaching me to back myself. Everyone should be so lucky as to have a PhD friend like you.

I am grateful for the excellent scientific advice of my thesis committee; thank you to Kathy Niakan, Rebecca Oakey, Mike Gilchrist, Greg Elgar, and Pontus Skoglund. Thanks also to members of the Niakan and Lovell-Badge labs, for scientific advice and a friendly community. I was also lucky to have support and advice from members of the Reik lab, most notably Stephen Clark.

I am very grateful to have had the support of several excellent Crick core facilities over the course of this project. Over and above the scientific support they have offered, I'm thankful for the camaraderie and conversation. Thank you to members of the ASF, especially Robert and Deb who gave me so much help and freedom to endlessly optimise weird and wonderful bisulfite libraries. To Stuart from BABs, who helped kick-start this

project. And to everyone in BRF Experimental 2, especially Jamie D, Nick, Jamie B, and Matt. Thank you for looking after the opossums so well.

Julia, the Horsfield lab and Otago Genetics – my first scientific home. Julia, when I was ruminating on career decisions, you wisely advised me not to 'differentiate too early'. It was your advice that committed me to undertaking a PhD. Thank you for giving me such a strong foundation.

My involvement with the Crick Netball club has been a wonderful Monday night distraction – thanks to you all, but especially Abi, Liz, and everyone from team Pass the Possum. Tuesday evening netball games with the Asteroids have also been a welcome break, as have Saturdays out in nature with Wild About Hampstead Heath.

I moved across the world to undertake this project, and yet I found a village of support. Thank you to Rosemary, for encouraging me even before I began, and together with John being my UK home. Thank you to Joey, Mel, Katy and the extended Hewett clan for lots of lovely Christmases, family visits, and muddy walks. I'm grateful to Granpa and Granny Hewett for always supporting my education.

Ally, Rachel, Michelle, Marina, Kate, Karla, Elsie, and Amarni; thank you for maintaining (with the help of Skype) our friendships across the sea. Sissy, who I met as a science friend, but I now count as a life-long one; thank you for dragging me along on many wild adventures. To Jordy, Max, and Berry, the three best flatmates I could've asked for, thank you for running London's finest Kiwi hotel with me. Charlotte, who is in here twice (!); thank you for organising so much fun, and for welcoming me so generously into the circle of your friends and family. Shaun, thank you for memorising key words from the epigenetics Wikipedia page in order to explain my job to your friends. You have variously provided distraction, consolation, and celebration when they were each needed, and always a sense of perspective. Thank you always.

Finally, to Mum, Dad, and Harrie. Thank you for instilling so many of the values that made me. Mum, it is because of you that I read voraciously, that I love nature, and that I know the strength that comes from talking things through. Dad, you taught me to be sceptical, not to shy away from hard challenges, and that everything can be a puzzle that is fun to solve. Harrie, the way you approach any challenge with unbridled and positive energy inspires me to do the same. Thank you, family, for your love and support. I dedicate this thesis to you.

Table of Contents

Table of figures

List of tables

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Bp	Basepairs
BRF	Biological Research Facility
CDS	Coding sequence
CGI	CpG island
Chromatin immunoprecipitation	ChIP
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
E	Embryonic day
EB	Elution Buffer
EGA	Embryonic genome activation
EtBr	Ethidium Bromide
FPKM	Fragments Per Kilobase per Million mapped reads
Gb	Gigabases
gDNA	Genomic DNA
GTF	Gene transfer format
ICR	Imprinting control region
Kb	Kilobases
Mb	Megabases
Mya	Million years ago
MZT	Maternal-zygotic transition
NEB	New England Biolabs
PCR	Polymerase chain reaction
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNA FISH	RNA fluorescence in situ hybridisation
Rpm	Rotations per minute
RRBS	Reduced-representation bisulfite sequencing
RT-PCR	Reverse-transcriptase polymerase chain reaction
scRNA-seq	Single cell RNA-seq
SNP	Single-nucleotide polymorphism
SNP	Single nucleotide polymorphism
TAE	Tris-acetate-EDTA
TSS	Transcriptional start site
TSS	Transcriptional start site
UCSC	University of California Santa Cruz
V	Volts
v/v	Volume/volume

w/v	Weight/volume
WGS	Whole-genome sequencing

Chapter 1. Introduction

The epigenetic control of gene expression is crucial to successful embryonic development, allowing different programmes of gene expression to emerge from cells descended from the totipotent zygote. In order to understand how epigenetic regulation of early development has evolved, it is crucial to examine the dynamics of epigenetic landscapes in the context of diverse organisms and life-history strategies. In this thesis, I focus on how the role of one key component of epigenetic regulation, DNA methylation, has evolved in the mammalian lineage.

1.1 Evolution and early development of mammals

Life history and reproductive strategies vary between different vertebrate lineages. In aquatic vertebrates such as the zebrafish, embryos develop unprotected from the environment and predation. Thus hundreds of zygotes are produced at spawning, and development proceeds quickly, resulting in free-swimming larvae within 48-72 hours of fertilisation (Kimmel et al., 1995). In terrestrial vertebrates, the egg evolved to have a thick shell to prevent desiccation, and extra-embryonic membranes to facilitate gas exchange and storage of waste products (Frankenberg, 2018). In addition, in order for the newly-hatched animal to successfully feed on land, a mature and motile body plan must be achieved by the time of hatching; therefore the eggs of terrestrial vertebrates evolved to contain a large volume of yolk to sustain an extended developmental period (Frankenberg, 2018).

Mammals are distinguished from other vertebrate animals by the presence of lactation as a means of post-natal maternal provision of nutrition to offspring (Figure 1), resulting in progeny that are reliant on maternal care immediately post-birth. The mammalian clade constitutes three groups; prototherians (monotremes), metatherians (marsupials), and eutherians (Figure 1). Prototherians are egg-laying mammals that diverged from the therian clade approximately 166 million years ago (mya) (Bininda-Emonds et al., 2007); extant members of this clade are the platypus and the echidna. Unlike prototherians, the therian mammals do not lay eggs; rather they have evolved a viviparous mode of

reproduction wherein the egg is produced with relatively less yolk, and the developing embryo is supported by maternally-supplied nutrition via the placenta (Frankenberg, 2018). The two therian clades diverged 160 mya (Luo et al., 2011), and share key mammalian features such as XY sex chromosomes and genomic imprinting in addition to viviparity (Figure 1). The relative allocation between pre- and post-natal phases differs between eutherians and metatherians. In metatherians, the gestation period is relatively short, and the period of placental attachment is brief (Tyndale-Biscoe and Renfree, 1987). At birth, highly altricial young crawl to the pouch, where the remainder of the developmental period is completed sustained by the mother's milk. For example, in the South American opossum *Monodelphis domestica* implantation occurs at 12.5 days gestation, followed by birth two days later (Mate et al., 1994). After birth, young continue to develop during a suckling period of at least two months before weaning occurs (Bergallo and Cerqueira, 1994). By contrast, in eutherians the relative period of placenta:uterine contact is considerably longer; for example, implantation occurs at embryonic day 4.5 (E4.5) in the mouse, with birth occurring at 19 days. The evolution of lactation and subsequently viviparity in therian mammals can be viewed as the adoption of a successful reproductive strategy in this lineage (Wourms and Callard, 1992).

In addition to differences in overall reproductive strategy, there are notable differences in the early development and embryology of different vertebrate groups (Figure 2). During early development, the embryo must undergo many cell divisions and organise these cells to form the different structures of the body plan. A key occurrence during this process is the maternal to zygotic transition in control of development (MZT) (Lee et al., 2014a). The oocyte is laden with mRNAs and proteins deposited by the mother during oocyte growth. At MZT, maternal products are degraded, and embryonic genome activation (EGA) occurs. This process is key to the ability of the embryo to achieve totipotency from the baseline established by the terminally differentiated gametes, and subsequently to achieve differentiation, by allowing different transcriptional programmes to occur in the different cell lineages (Lee et al., 2014a).

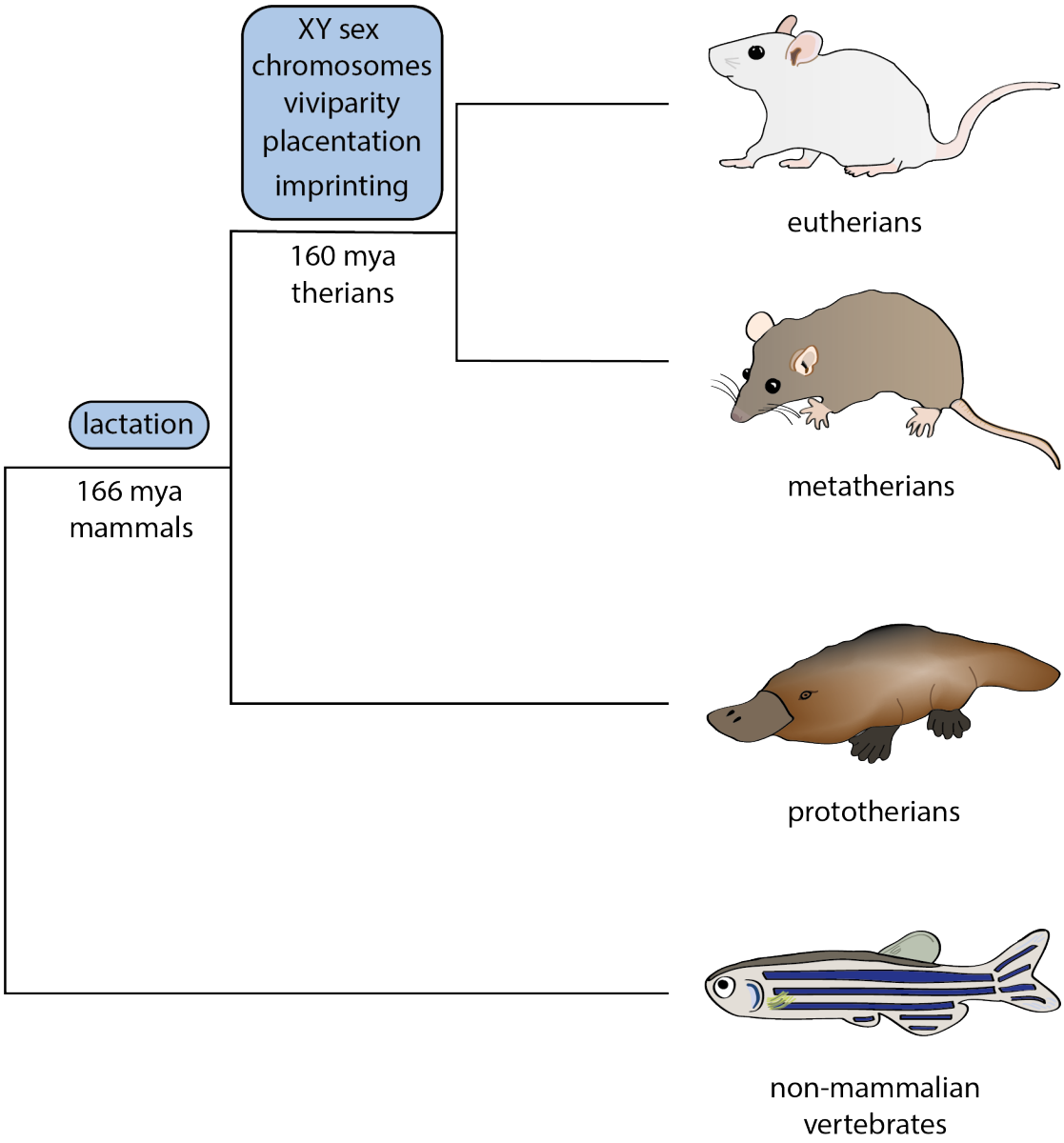


Figure 1: Evolution of mammals. Schematic representing the evolutionary relationship between different mammalian groups, with non-mammalian vertebrates as an out-group. Evolutionary innovations in each mammalian lineage are indicated in blue boxes. mya: million years ago.

In the zebrafish, the maternal products sustain the embryo throughout the period of rapid cleavage divisions (Kane and Kimmel, 1993; Figure 2a). After 10 cell cycles, developmental control is passed from the mother to the embryo at EGA (Kane and Kimmel, 1993, Kimmel et al., 1995; Figure 2a). The factors that control EGA are well-studied in non-mammals such as the zebrafish (Lee et al., 2014a). In the zebrafish, the maternally deposited transcription factors Pou5f1, Nanog, and Sox19b are important activators of zygotic transcription (Lee et al., 2013). One of the targets of this activation

is the microRNA *miR-430*, the expression of which results in the degradation of maternal transcripts (Giraldez et al., 2006). Notably, removal of Pou5f1, Nanog, and Sox19b resulted in a phenotype resembling alpha-amanitin treatment (Lee et al., 2013). Shortly after EGA, the formation of the somatic cell lineages occurs at gastrulation (Figure 2a). Zebrafish embryos in which transcription was inhibited by injection of alpha-amanitin were shown to proceed normally through the cleavage and early blastula phases (Kane et al., 1996). However, inhibition of transcription caused developmental arrest just prior to the onset of gastrulation, indicating that in zebrafish, it is at gastrulation that embryonic transcripts are first required (Kane et al., 1996).

In eutherians, the early embryo also passes through a series of cleavage divisions (Figure 2b). However, in comparison to the zebrafish, EGA occurs after far fewer cell divisions. In the mouse, major EGA occurs at the two-cell stage, and in human embryos at the four-eight cell stage (Golbus et al., 1973, Warner and Versteegh, 1974, Hamatani et al., 2004, Dobson et al., 2004, Vassena et al., 2011, Niakan et al., 2012). In both the mouse and the human embryo, minor waves of transcription can also be detected at earlier stages (Niakan et al., 2012). Mouse embryos cultured in alpha-amanitin arrest development at the two-cell stage, implying that the first requirement for transcription from the embryonic genome is at the major genome activation (Golbus et al., 1973, Warner and Versteegh, 1974). Similarly, in human embryos transcription inhibition results in developmental arrest at the four-eight cell stage (Braude et al., 1988).

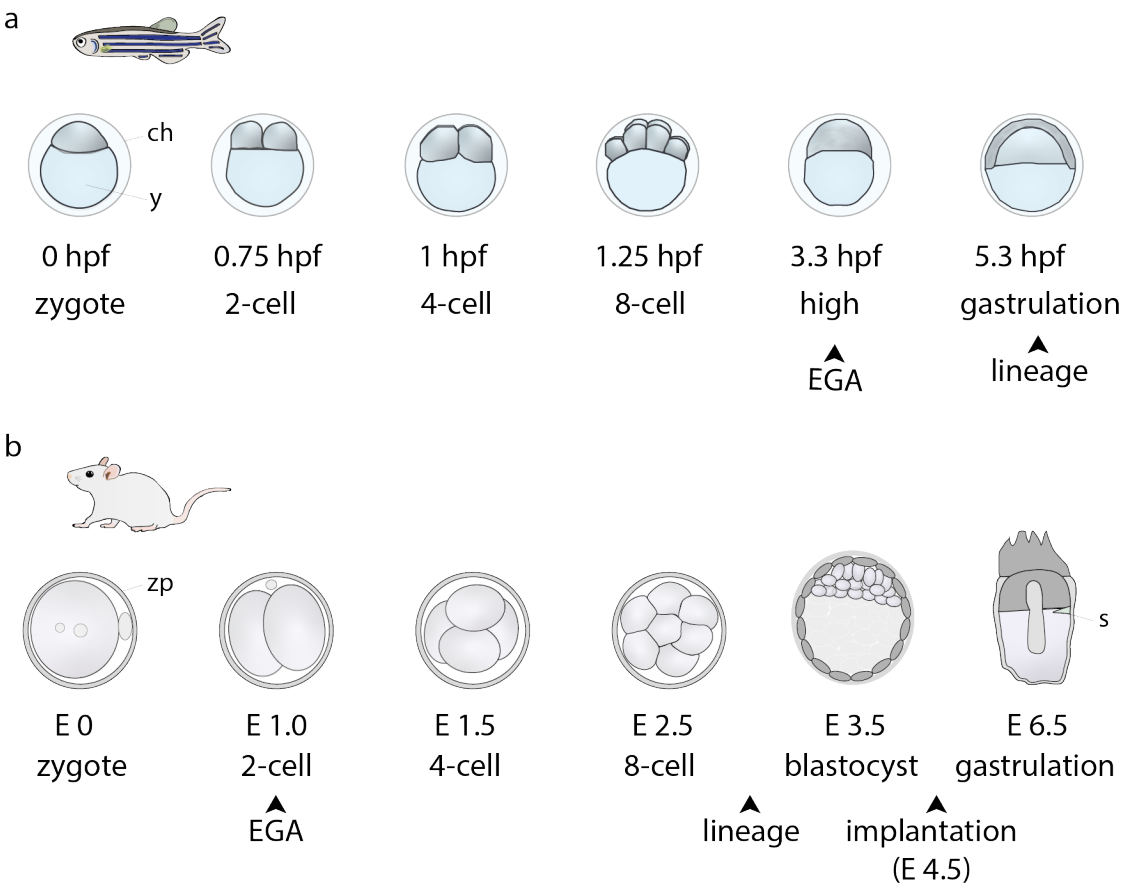

a
ch
y
0 hpf
zygote
0.75 hpf
2-cell
1 hpf
4-cell
1.25 hpf
8-cell
3.3 hpf
high
EGA
5.3 hpf
gastrulation
lineage
b
zp
E 0
zygote
E 1.0
2-cell
EGA
E 1.5
4-cell
E 2.5
8-cell
lineage
E 3.5
blastocyst
implantation
(E 4.5)
s
E 6.5
gastrulation
c
sh
m
zp
y
E0-1
zygote
E1.5
2-cell
E2.5
4-cell
E3.5
8-cell
EGA
E6 - 6.5
blastocyst
lineage
s
E8.5-9
gastrulation
implantation
(12.5 dpc)

Figure 2: Embryonic development of the zebrafish, mouse, and opossum. Schematic illustrating key developmental stages in embryos of the zebrafish, mouse, and opossum. **a**: Zebrafish development initially consists of rapid cleavage divisions, followed by the transition to embryonic control of development with embryonic genome activation (EGA) from the tenth cell cycle at the 'high' stage 3.3 hours post fertilisation (hpf). EGA is soon followed by gastrulation, resulting in the formation of different cell lineages in the form of the three germ layers. **b**: In the mouse, major EGA occurs at the two-cell stage. Mouse development consists of cleavage divisions until the eight-cell stage, at which time compaction occurs to form a morula. By E3.5, different cell lineages have been a specified and blastocyst has formed, consisting of a pluripotent inner cell mass surrounded by extra-embryonic trophectoderm cells. Implantation occurs at E4.5, followed by gastrulation from E6.5. **c**: In opossums, the embryo is enclosed within a zona pellucida, a thick glycoprotein mucoid coat, and a thin outer shell coat. The cleavage divisions are accompanied by extrusion of yolk mass/deutoplasm from the blastomeres into the space enclosed by the zona pellucida. No blastomere-blastomere adhesions are formed and the embryo does not pass through a morula stage. Upon the formation of a blastocyst, blastomeres adhere to the zona pellucida. In the opossum, EGA occurs at E3.5. Differentiation of embryonic and extra-embryonic lineages occurs at day six, and formation of the germ layers occurs in the gastrulating embryo from day 8.5. The gastrulating opossum embryo is shown from a top view. ch: chorion, zp: zona pellucida, m: mucoid coat, y: yolk mass/ deutoplasm, s; primitive streak.

The transcriptional programme of the mammalian early embryo is characterised by the expression of transposons, for example MuERV-L and IAPs in the mouse. In addition, many protein-coding genes are expressed from transposon-derived promoters and enhancers (Kigami et al., 2003, Peaston et al., 2004, Macfarlan et al., 2012, Brind'Amour et al., 2018). Recent studies have investigated the control of EGA in eutherians. In the mouse, transcription factors of the *Zscan4* cluster become expressed in the two-cell embryo, and regulate cleavage stage genes, including endogenous retroviruses (Falco et al., 2007, Zhang et al., 2019). In both humans and mice, DUX family transcription factors were observed to bind the promoters of genes activated early in EGA, along with several groups of transposable elements known to gain expression at EGA (Macfarlan et al., 2011, Macfarlan et al., 2012, Hendrickson et al., 2017, De Iaco et al., 2017). The developmental pluripotency associated factors DPPA2 and DPPA4 were shown to act upstream of DUX and ZSCAN4 in regulating EGA genes (Eckersley-Maslin et al., 2019). However, while *Dux* null embryos displayed aberrant EGA and a range of developmental defects, some *Dux* null animals were able to successfully complete development, which suggests that other factors redundantly regulate EGA in eutherians (Chen and Zhang, 2019, Guo et al., 2019, De Iaco et al., 2020).

Following EGA and the cleavage stages, the eutherian embryo undergoes compaction, forming cell-cell adhesions to generate a morula. In the mouse, the blastocyst is formed by E3.5 (Figure 2b), and implantation occurs one day later, at E4.5. The eutherian pre-implantation blastocyst comprises the pluripotent inner cell mass (ICM), and the trophectoderm (TE). The ICM contributes to both the epiblast, which forms the embryonic lineage, and the primitive endoderm (PrE), which will form the yolk sac. The TE will form the trophoblast cells of the placenta. Differentiation of the germ layers at gastrulation occurs afterwards, at E6.5 in the mouse (Figure 2b). Therefore, unlike in non-mammalian vertebrates, the first lineage divergence in eutherians is the specification of the extra-embryonic lineages. The occurrence of eutherian EGA far in advance of gastrulation is likely due to this uniquely mammalian requirement to specify the extra-embryonic cell lineages.

Metatherian preimplantation embryos complete a programme of development similar to eutherians in some respects and differing in others (Figure 2c). The metatherian embryo is enclosed within a zona pellucida, as in eutherians, but also within a thick glycoprotein mucoid layer and a thin outer shell coat (Selwood, 2000). The outer coats are deposited by the maternal cells of the reproductive tract as the embryo transits from the ovary to the uterus and may act as barrier to polyspermy. The mucoid coat becomes thinner as development progresses, and though the mechanism of its breakdown is not known, it has also been suggested to provide nutrients to the developing embryo (Arnold and Shorey, 1985b, Arnold and Shorey, 1985a, Selwood, 2000; Figure 2c). After fertilisation, a polarised zygote is formed with transparent yellow cytoplasm at the upper pole and opaque grey cytoplasm at the lower pole. Polarisation is maintained during subsequent cleavage divisions, with blastomeres loosely grouping at the upper pole. The extrusion of cytoplasmic material from the blastomeres forms a ‘yolk mass’, or deutoplasm, towards the lower pole (Selwood, 1992, Selwood et al., 1997, Mate et al., 1994; Figure 2c). During the cleavage divisions there are no cell adhesions between blastomeres and consequently metatherian embryos do not form a morula.

Dividing blastomeres eventually ring the zona pellucida and adhere to it, forming a blastocyst. The very early blastocyst is composed of visually uniform cells; subsequently two cell types become morphologically discernible, with the embryonic compartment forming smaller, rounded cells and the TE forming larger flattened cells (Selwood et al., 1997, Selwood, 1992, Frankenberg et al., 2013, Mate et al., 1994; Figure 2c). Unlike eutherians, the metatherian blastocyst has no ICM. A bilaminar blastocyst forms by the inward migration of cells adjacent to the embryonic area (Selwood and Johnson, 2006, Mate et al., 1994). The formation of the primitive streak at gastrulation results in a trilaminar blastocyst, and subsequently the embryo rapidly completes organogenesis (Mate et al., 1994).

Relatively little is known about EGA in metatherian species, in part due to the absence of robust embryo culture approaches precluding transcription inhibition experiments. However, a recent study undertook transcriptomics analysis of the early embryos of the opossum *M. domestica* and revealed that EGA occurs at E3.5 in this species (approximately the eight-cell stage; Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished). In dimensionality reduction analysis of transcripts, oocytes through to E2.5 embryos grouped together, indicating their transcriptomes were similar, while embryos from E3.5 onwards clustered by timepoint. In addition, transcripts from the Y chromosome, which cannot be maternally deposited, increased between E2.5 – 3.5. Metatherian EGA therefore occurs at an earlier cell stage than in non-mammalian animals like the zebrafish. The earlier timing of EGA may be required to facilitate the divergence between embryonic and extraembryonic lineages, similar to eutherian mammals.

The complex processes of early development result in the differentiation of multiple cell lineages under the control of the zygotic genome. Transcriptional control over development is accompanied by a suite of epigenetic regulatory mechanisms that aid the zygotic genome in generating the transcriptional programs necessary for successful development.

1.2 Epigenetics and the regulation of transcription

By differentially regulating gene expression between cells within the same organism, different cell states can be achieved from the same underlying genome; this phenomenon underlies the complex developmental trajectories of multicellular organisms. Complex networks of transcription factors are instructive in establishing gene-expression patterns, but they do not work in isolation; a suite of epigenetic information influences, and is influenced by, the transcriptional machinery to result in each particular cell state. The meaning of the term epigenetics has evolved over time (Greally, 2018), but is commonly used to refer to the layers of molecular information that exist over and above the DNA sequence, and that are associated with the regulation of gene expression. The term epigenetics has also been used to mean the retention of such information as 'cellular memory' either through mitotic cell divisions or through the germline (Deichmann, 2016). These layers of molecular information comprise several forms; direct modifications of DNA, such as DNA methylation; post-translational modifications of the histones around which the DNA is wound; the actions of non-coding RNAs; and the overall structure of the chromatin, both at local scales and the large-scale three-dimensional organisation of the genome. These epigenetic mechanisms do not act in isolation, but are fundamentally intertwined and can promote or antagonise each other. Given the focus upon DNA methylation in this thesis, in the subsequent sections I will briefly outline other forms of epigenetic regulation before exploring DNA methylation in greater detail.

1.2.1 **Post-translational modification of histones**

DNA is packaged into chromatin by its winding around the histones of the nucleosome (Kornberg and Thomas, 1974, Luger et al., 1997). The tails of histones are subject to many post-translational modifications, which are dynamically deposited and removed by a suite of histone-modifying enzymes (Bannister and Kouzarides, 2011). Histone modifications can influence the structure of chromatin and can also affect the binding of effector molecules and are therefore suggested to act as functional regulators of the genome; for example, by creating or blocking accessibility of chromatin to factors involved in transcription (Strahl and Allis, 2000, Turner, 2000, Jenuwein and Allis, 2001). Indeed, different histone modifications are associated with different functional

genomic regions. For example, acetylation of histones neutralises the charged histone tails, and renders chromatin more accessible; acetylation is enriched at promoters and enhancers (Wang et al., 2008). Active promoters are decorated with H3K4me3 (Schneider et al., 2004), and actively transcribed genic regions are enriched for H3K36me3 (Bannister et al., 2005). Conversely, silenced regions of the genome are also decorated with particular histone modifications; H3K27me3 is enriched at facultative heterochromatin, and H3K9me is enriched at constitutive heterochromatin (Peters et al., 2003). Promoters harbouring both H3K4me3 and H3K27me3 are thought to be held in a poised state, to be activated later, for example during differentiation (Guenther et al., 2007, Mikkelsen et al., 2007a, Bernstein et al., 2006).

1.2.2 **Non-coding RNAs**

Much of the genome is transcribed but does not code for proteins (Encode Project Consortium, 2012), instead producing a variety of small RNAs and long non-coding RNAs (lncRNAs), both of which are involved in transcriptional regulation by acting as scaffolds to recruit chromatin modifiers and thereby altering epigenetic state (Holoch and Moazed, 2015). In various organisms, small RNAs can induce silencing via RNA interference (RNAi) pathways (Fire et al., 1998), via degradation of transcripts complementary to the small RNAs (Hammond et al., 2000), and via the induction of repressive histone and DNA modifications in the nucleus, for example silencing of transposons via H3K9me3 in the *Drosophila* germline (Sienski et al., 2012), and via DNA methylation in the mouse germline (Aravin et al., 2008). lncRNAs are also involved in the regulation of transcription, by recruiting and targeting chromatin-modification complexes (Khalil et al., 2009). For example, many human lncRNAs were found to interact with the polycomb repressor complex 2 (PRC2), and at some loci repressive H3K27me3 was shown to depend on the expression of lncRNAs (Khalil et al., 2009). In another example, a class of lncRNAs termed eRNAs are transcribed from enhancer sequences (Kim et al., 2010, De Santa et al., 2010), and are thought to act by recruiting transcriptional activators and facilitating their interaction with the promoter via 3D chromatin looping (Lai et al., 2013).

1.2.3 **Chromatin accessibility and three-dimensional genome architecture**

The accessibility of chromatin, as measured by vulnerability to the action of DNA-targeting enzymes in assays such as DNase I hypersensitivity and ATAC-seq, reflects the regulatory status of the genome (Gross and Garrard, 1988, Felsenfeld et al., 1996, Thurman et al., 2012, Buenrostro et al., 2015). At regulatory regions such as enhancers, promoters, and insulators, the binding of transcription regulators displaces nucleosomes that would otherwise be present, resulting in nucleosome-depleted accessible sites (Thurman et al., 2012). By contrast, heterochromatic regions are densely occupied by nucleosomes.

At a higher level, the genome assumes a complex condensed 3D structure. This condensed structure is reflective of the need to efficiently package the genome inside the nucleus, and yet is also correlated with genome activity. In the interphase nucleus, visualisation of entire chromosomes by fluorescence *in situ* hybridisation revealed that chromosomes tend to occupy discrete territories, with gene-poor chromosomes more likely to be situated at the nuclear periphery, and gene-rich chromosomes the nuclear interior (Tanabe et al., 2002, Bolzer et al., 2005, Cremer and Cremer, 2010). The development of chromosome conformation capture techniques to map interacting regions of the genome revealed that the genome adopts complex 3D conformations at various scales (Dekker et al., 2002, Lieberman-Aiden et al., 2009, Rao et al., 2014). At the local scale, chromatin forms loops that bring together regulatory regions and genes that are separate in linear genomic space. At a larger scale, DNA forms regions of increased interaction frequency termed topologically-associated domains (TADs). At a global scale, the genome is divided into active and inactive regions termed A and B compartments. Increasing resolution of Hi-C studies has revealed that the scale at which these conformations are perceived to occur can be a function of the genomic resolution attained by the experiment (Rao et al., 2017), and because conformation capture methods assess a population of cells, these 3D structures represent a population average rather than a static state present in every cell. Architectural proteins such as CCCTC-binding factor (CTCF) and the cohesin complex play an important role in the organisation of chromatin structure (de Wit et al., 2015, Rao et al., 2017).

1.3 DNA methylation

It has long been known that DNA contains modified bases, the predominant one being methylation at 5' carbon of cytosine, typically in CpG context (Grippo et al., 1968).
DNA methylation was proposed as a putative stable epigenetic mark due to the palindromic nature of the CpG site (Riggs, 1975, Holliday and Pugh, 1975). This hypothesis relied on the idea, unproven at the time, of an enzyme that could recognise hemi-methylated DNA after replication and faithfully copy it to the daughter strand, thereby perpetuating methylation patterns through cell division. Such a mechanism was subsequently verified (Bird, 1978, Pollack et al., 1980, Wigler et al., 1981). Methylation was viewed as a repressive mark responsible for gene silencing, based on early experiments that demonstrated lower methylation in active genomic regions and a repressive effect of methylation at promoters (Naveh-Many and Cedar, 1981, Ben-Hattar and Jiricny, 1988). Subsequently, methylation was found to be enriched at silenced promoters on the mammalian inactive X (Xi), to have a role in controlling the allele-specific expression of imprinted genes, and to prevent the expression of transposable elements (Mohandas et al., 1981, Ferguson-Smith et al., 1993, Li et al., 1993, Walsh et al., 1998). The importance of this epigenetic mark was underscored by the requirement for DNA methylation in order to successfully complete embryonic development (Li et al., 1992b, Okano et al., 1999).

1.3.1 **The DNA methylation machinery**

De novo DNA methylation is established by the DNA methyltransferase (DNMT) enzymes DNMT3A and DNMT3B (Okano et al., 1999). The *de novo* methyltransferases do not appear to possess sequence specificity (Okano et al., 1998, Dodge et al., 2002), and therefore cross-talk with other proteins and epigenetic marks serve to direct their pattern of methylation establishment. For example, H3K4 methylation at CpG island (CGI) promoters inhibits DNMT3 enzymes, protecting hypomethylated CGIs from *de novo* methylation (Otani et al., 2009). Conversely, H3K9 methylation and DNA methylation are strongly correlated. Some SET-domain enzymes responsible for H3K9 methylation are able to recruit DNMT3 enzymes, resulting in directed *de novo* methylation (Du et al., 2015). DNMT3L, a catalytically inactive DNMT, co-operates with

de novo methyltransferases in the correct establishment of methylation in the germline and in the early embryo (Bourc'his et al., 2001, Hata et al., 2002, Webster et al., 2005, Guenatri et al., 2013). Orthologues of DNMT3A and DNMT3B are present in non-mammalian vertebrates such as fish, while DNMT3L evolved specifically in the therian lineage, and may therefore have arisen together with genomic imprinting (Yokomine et al., 2006). In addition, a rodent-specific methyltransferase, DNMT3C, is expressed in the male germline where it silences transposons to ensure successful meiosis (Barau et al., 2016).

Existing patterns of DNA methylation are maintained during cell division as a function of the palindromic nature of CpG sites. The maintenance methyltransferase DNMT1 and its binding partner UHRF1 localise at the replication fork and recognise the hemi-methylated DNA generated by DNA replication (Sharif et al., 2007, Bostick et al., 2007, Song et al., 2012). DNMT1 then faithfully copies the methylation pattern of the parent strand onto the daughter strand. This mechanism ensures the heritability of DNA methylation through cell divisions.

Removal of methylation can be achieved via a passive dilution through cell divisions due to the absence of maintenance methyltransferase activity. Active removal of DNA methylation can also occur via the action of the ten-eleven translocation enzymes (TET1, 2, and 3). TET enzymes oxidise methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Tahiliani et al., 2009, He et al., 2011, Ito et al., 2011). The oxidised bases can be lost passively due to inefficient recognition by maintenance methylation at replication, or by excision of the base by thymine-DNA glycosylase (TDG), and subsequent base excision repair (BER; He et al., 2011).

Methylation state can be ‘read’ by DNA-binding proteins that maintain a preference for methylated, or unmethylated DNA (Ren et al., 2018). Proteins with such functions have been identified, including the methyl-CpG binding domain (MBD) family of proteins (Meehan et al., 1989, Lewis et al., 1992), as well as several zinc finger proteins and transcription factors (Sasai et al., 2010). The MBD family includes proteins MBD1-4 and MeCP2, of which MBD1, 2, 4, and MeCP2 bind to methylated DNA via the MBD

(Baubec et al., 2013). MBD3 does not preferentially bind methylated DNA (Saito and Ishikawa, 2002, Baubec et al., 2013). MBD proteins interact with histone deacetylase complexes, and are thereby involved in establishment of repressive chromatin at sites of DNA methylation, though domain knockouts suggested that other regions of MBD proteins can bind unmethylated DNA (Nan et al., 1998, Ng et al., 1999). *In vitro* screens for transcription factor binding to methylated and unmethylated DNA templates identified instances where methylation either promoted or deterred transcription factor binding (Kribelbauer et al., 2017, Yin et al., 2017).

1.3.2 **Distribution and roles of DNA methylation across the genome**

DNA methylation is found in many eukaryotes, and genome-wide patterns differ between groups. In plants, DNA methylation is found on cytosine in CpG, CHG (H = A, C, or T), and CHH contexts. Heavy methylation is found at transposable elements and some euchromatic loci, and bodies of active genes are methylated (Cokus et al., 2008, Lister et al., 2008). The distribution can be mosaic, with interspersed regions of high and low methylation, or global, depending on species (Suzuki and Bird, 2008). DNA methylation is absent in some insect species, but present in others (Bewick et al., 2017). In those insects where DNA methylation is present, it is mostly found in the CpG context, though fewer than 15% of CpGs carry methylation, and most methylation is located in the transcribed regions of genes (Bewick et al., 2017, Rehan et al., 2016, Glastad et al., 2016).

In vertebrate genomes, methylation is most prevalent at cytosines in the CpG context, although methylation of non-CpG cytosines is also observed in embryonic stem cells, developing neurons, and oocytes, albeit at much lower frequency than CpG methylation (Lister et al., 2009a, Tomizawa et al., 2011, Lister et al., 2013). Vertebrate genomes are largely hypermethylated, with methylation in a bimodal distribution where most CpG sites are highly methylated and a smaller number are unmethylated (Meissner et al., 2008; Figure 3a). CpG sites are underrepresented in vertebrate genomes due to the propensity for methylated cytosine to spontaneously deaminate to thymine (Bird, 1980, Shen et al., 1994); the exceptions to this rule are regions of densely clustered CpGs termed CpG islands (CGIs), which are typically unmethylated and are therefore protected from

deamination (Cooper et al., 1983, Bird et al., 1985, Gardiner-Garden and Frommer, 1987).

CGIs very frequently overlap with gene promoters; in particular, most widely-expressed 'housekeeping' genes possess a CpG island (Gardiner-Garden and Frommer, 1987, Larsen et al., 1992). Gene repression as a result of hypermethylation of CGI promoters has been the classical example of gene regulation by DNA methylation (McGhee and Ginder, 1979), but in fact methylation at most CGIs does not vary between tissues; CGIs are almost invariably unmethylated (Figure 3b). Exceptions where promoter CGI methylation does occur include the promoters of silenced genes on the inactive X chromosome in female mammals, and of germline genes in the vertebrate soma (Borgel et al., 2010). In contrast, the regions immediately adjacent to CGIs, termed CpG island shores, were observed to vary in their methylation level between tissues, a property that correlates with the expression level of nearby genes (Irizarry et al., 2009; Figure 3b).

In fact, it is at distal regulatory elements such as enhancers where methylation varies most between tissues (Stadler et al., 2011, Hodges et al., 2011, Bock et al., 2012). At enhancers, DNA methylation level is low, though not entirely absent (Stadler et al., 2011; Figure 3c). Lower methylation correlates with enhancer activity, while poised or silent enhancers harboured higher methylation (Stadler et al., 2011, Hodges et al., 2011, Bock et al., 2012; Figure 3c). In the cases of CCTC-binding factor (CTCF) and the transcription factor REST, decreased methylation at enhancers was shown to occur as result of transcription factor binding at the locus. The presence of 5hmC at these loci suggested the involvement of active demethylation (Stadler et al., 2011, Feldmann et al., 2013). However, in an *in vitro* survey of 542 human transcription factors, methylation of DNA prevented binding of some proteins, promoted binding for others, and in some cases transcription factors were agnostic to methylation state; therefore the influence of methylation on transcription factor activity is nuanced (Yin et al., 2017). A recent study examined the temporal relationship between chromatin accessibility and DNA methylation of regulatory regions in a model of cell fate transition (Barnett et al., 2020). Opening of the chromatin and resultant increases in the expression of nearby genes preceded changes in DNA

methylation, implying that DNA methylation states at distal regulatory regions are reflective of transcription factor binding rather than instructive for it (Barnett et al., 2020).

In addition to CpG islands and distal regulatory regions, vertebrate genomes were also observed to contain large stretches (3.5-25 kb) of unmethylated DNA, termed DNA methylation canyons (Xie et al., 2013, Jeong et al., 2014; Figure 3d). Most canyons contained a CGI or promoter/genic regions, but were larger and comprised a lower GC content than canonical CGIs (Xie et al., 2013, Jeong et al., 2014). Canyons were particularly enriched for genes involved in development and lineage markers, but were present in adult tissues as well as developmental cell types (Xie et al., 2013, Jeong et al., 2014). Overlap between canyons and highly conserved non-coding regions suggested these regions represent cis-regulatory elements (Jeong et al., 2014). The demarcation of methylation canyons was found to be dependent on a interplay between DNMT3a and TET proteins (Jeong et al., 2014, Wiehle et al., 2016). Aside from regions of low or variable methylation, the remainder of vertebrate genomes are highly methylated, including at gene bodies and repetitive elements.

Gene bodies harbour high levels of DNA methylation, a pattern that is conserved across many groups, including vertebrates (Hellman and Chess, 2007, Lister et al., 2009b), invertebrates (Sarda et al., 2012, Suzuki et al., 2007a), and plants (Zilberman et al., 2007; Figure 3e). Gene body methylation is associated with active transcription, a finding seemingly at odds with the traditional view of DNA methylation as a mark of silenced DNA. This pattern results from the cross-talk between histone modifications and the *de novo* methylation machinery; H3K6me3 is deposited alongside active transcription, and is recognised by the PWWP domain of DNMT3B, bringing *de novo* methylation to transcribed regions (Baubec et al., 2015, Morselli et al., 2015). An increase in promiscuous transcripts originating from non-canonical intragenic promoters was observed in *Dnmt3b*-null embryonic stem cells, suggesting that one role for gene body DNA methylation may be the suppression of cryptic promoters that might otherwise be available to RNA polymerase II entry due to the open chromatin structure of transcribed regions (Neri et al., 2017, Teissandier and Bourc'his, 2017). The observation that exons appear more highly methylated than introns suggested that genic DNA methylation might

also play a role in regulation of splicing (Lister et al., 2009b). In support of this notion, it was shown that alternatively spliced exons have a higher methylation level, and that loss of methylation dysregulated splicing events (Maunakea et al., 2013). Binding of MeCP2 was enriched at highly methylated exons, and positively regulated their inclusion during splicing (Maunakea et al., 2013).

A large proportion of vertebrate genomes is comprised of repetitive sequences, which must be controlled to avoid harm to the genome via transposition or recombination between non-allelic repeat elements. It was therefore proposed that extensive methylation in vertebrates evolved as a genomic defence system against repeat-induced damage (Yoder et al., 1997). Lack of methylation in early mouse embryos due to the removal of DNMT1 results in activation of intracisternal A particles (IAPs; Walsh et al., 1998). In the mouse, DNMT3L and the rodent-specific DNMT3C are required for the germline silencing of evolutionarily young transposons, which pose a threat to genome integrity due to their continued ability to mobilise (Bourc'his and Bestor, 2004, Barau et al., 2016). The presence of DNA methylation also prevents transposons from acting as aberrant sites of meiotic recombination in the germ cells of male mice (Zamudio et al., 2015). Satellite repeats are also subject to heavy methylation mediated by DNMT3B (Xu et al., 1999). Loss-of-function mutations in human DNMT3B cause ICF (immunodeficiency, centromere instability and facial anomalies) syndrome, in which the satellite repeats are completely unmethylated, resulting in chromosome instability (Xu et al., 1999). Therefore one of the functions of DNA methylation is to protect the genome from the potential negative consequences of housing repetitive DNA.

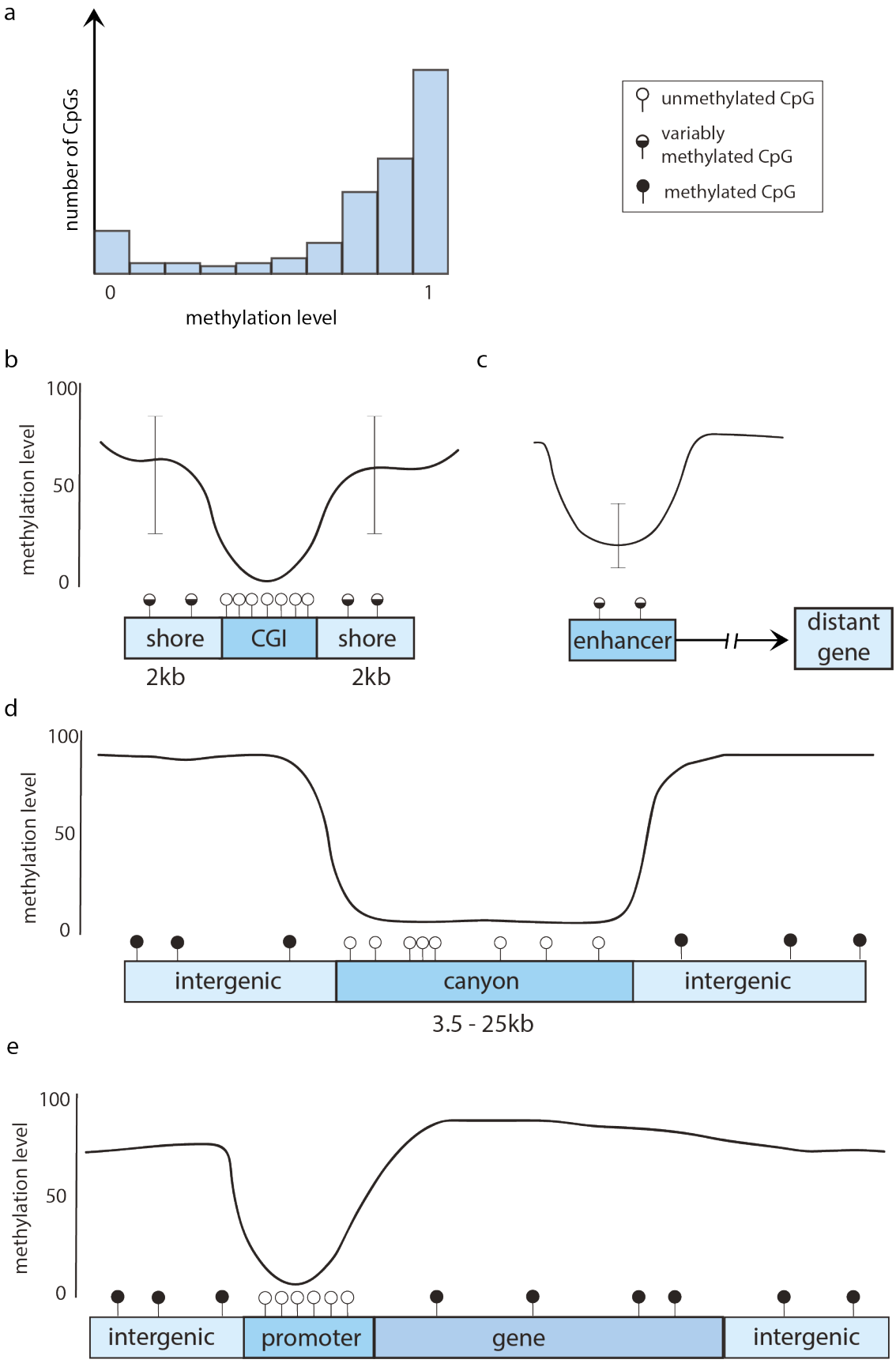

a
number of CpGs
0
1
methylation level
unmethylated CpG
variably methylated CpG
methylated CpG
b
methylation level
100
50
0
shore
CGI
shore
2kb
2kb
c
enhancer
distant gene
d
methylation level
100
50
0
intergenic
canyon
intergenic
3.5 - 25kb
e
methylation level
100
50
0
intergenic
promoter
gene
intergenic

Figure 3: Distribution of DNA methylation in vertebrate genomes. Schematic representations of the typical methylation distribution at some features of vertebrate genomes. **a**: Vertebrate genomes are hypermethylated in bimodal distribution, with most CpGs carrying high methylation, and a smaller number of CpGs carrying low methylation. **b**: Methylation level at a CGI and its shores. Methylation is low at most CGIs in the genome, but the regions approximately 2kb up- and downstream of CGIs, 'CGI shores', are often variably methylated between tissues. **c:** Methylation at enhancers varies depending on tissue type, with low methylation being correlated with enhancer activation. **d**: Methylation canyons are large domains (3.5 – 25kb) of low methylation. Some canyons encompass CGIs, but are larger than CGIs, and not as CG rich. Canyons are enriched for genes involved in development and markers of lineage and also overlap highly conserved regulatory regions. **e**: Typical methylation distribution over a gene and its promoter. Methylation is low at the promoter, which may overlap a CpG island (CGI). Methylation level over the gene body is high.

In summary, DNA methylation performs a variety of functions that are dependent on its genomic localisation and interplay with other epigenetic marks and DNA-binding factors. In some instances, changes to methylation state may be the result of transcriptional changes rather than the cause, but in other cases methylation clearly has a role in repressing transcription. It is clear that one function of methylation is to permanently silence loci in the context of cells where the factors that would permit activation of that locus are present; for example transposons, promoters of the inactive X chromosome in female mammals, and imprinted loci (Bestor et al., 2015). Lastly, while DNA methylation is largely static in the adult soma, in eutherian mammals there are dynamic changes during germline development and early embryogenesis. It is therefore interesting to consider the roles of methylation in the mammalian evolutionary lineage.

1.4 Genomic imprinting

Genomic imprinting, the phenomenon whereby alleles are differentially expressed in a parent-of-origin dependent manner, occurs in eutherian and metatherian mammals, and in flowering plants. Parent-of-origin specific expression was observed in maize in 1970, with the finding that a pigmentation gene was preferentially expressed from the maternal allele (Kermicle, 1970). The non-equivalence of the parental genomes of eutherian mammals was demonstrated by experiments in which gynogenetic or androgenetic embryos were shown to be non-viable (McGrath and Solter, 1983, Barton et al., 1984, Surani and Barton, 1983, Surani et al., 1984), followed by studies that identified the first

known mammalian imprinted loci, *Igf2r*, *Igf2*, and *H19* (Cattanach and Kirk, 1985, Barlow et al., 1991, Bartolomei et al., 1991, DeChiara et al., 1991, Ferguson-Smith et al., 1991).

In common between therian mammals and flowering plants is the existence of a system of maternal nutrient provision to the developing embryo, via the placenta in therian mammals and the endosperm in plants. One hypothesis to explain the evolution of imprinting states that in organisms in which maternal resource provision can be influenced by the genes of the developing embryo, a parental conflict arises (Moore and Haig, 1991). Paternal alleles are under selection to acquire as much maternal nutrition as possible, while it is in the mother's interest to withhold nutrients in anticipation of future offspring (Moore and Haig, 1991). In mouse, 197 imprinted genes are now reported, 63 of which overlap with the 165 reported imprinted genes in humans (Tucci et al., 2019). Many imprinted genes have roles in the placenta and foetal growth, while others are involved in postnatal behaviour (Tucci et al., 2019).

1.4.1 **Epigenetic regulation of imprints by DNA methylation**

The differential regulation of two alleles contained within the same nucleus is achieved via differential epigenetic marks established in the gametes. Following the identification of imprinted gene loci, differential methylation in male and female gametes was discovered at imprinted loci (Bartolomei et al., 1993, Ferguson-Smith et al., 1993, Li et al., 1993, Stoger et al., 1993).

In eutherians, imprinted genes largely occur in clusters, with allelic expression of the genes within the cluster controlled by differentially methylated regions (DMRs) between gametes, termed imprinting control regions (ICRs; (Edwards and Ferguson-Smith, 2007). Further imprinted DMRs (somatic DMRs) are established after fertilisation, but they are reliant on DMRs established in the germline (Edwards and Ferguson-Smith, 2007). ICR methylation is established in the gametes via the action of DNMT3A and DNMT3L (Bourc'his and Bestor, 2004, Bourc'his et al., 2001, Kaneda et al., 2004). After fertilisation, the zinc-finger proteins ZFP57 and ZNF445 interact with KRAB-associated protein 1 (KAP1) and the maintenance methylation proteins UHRF1 and DNMT1 to

maintain methylation at ICRs (Sharif et al., 2007, Hirasawa et al., 2008, Li et al., 2008, Quenneville et al., 2011, Takahashi et al., 2019). DNMT1 is found in two isoforms; the DNMT1S isoform is expressed at low levels in the oocyte and is also expressed by the embryo from the two-cell stage, and its protein is present in the nucleus throughout the cleavage stages, albeit at low amounts (Cirio et al., 2008b, Kurihara et al., 2008). DNMT1O, however, is produced at high concentration by the oocyte, and is carried through to the zygote, but is located in the cytoplasm apart from at the 8-cell stage when it translocates to the nucleus (Howell et al., 2001, Cirio et al., 2008b, Cirio et al., 2008a). Knockout of DNMT1O results in the dilution of methylation at imprints during exactly one cell-cycle, at the eight-cell stage (Howell et al., 2001, Toppings et al., 2008, Cirio et al., 2008a). This loss of methylation results in a mosaic of epigenotypes at imprinted loci, and subsequently an array of post-implantation developmental defects (Howell et al., 2001, Toppings et al., 2008, Cirio et al., 2008a, McGraw et al., 2013). In mice with a DNMT1 knockout that abolished both isoforms, all maintenance of methylation at imprints was lost, demonstrating that the DNMT1S isoform is responsible for maintenance of imprints during all other stages of the early embryo (Hirasawa et al., 2008).

It has been shown that STELLA, a maternal effect gene required for early development in the mouse, has a role in maintaining methylation at imprinted regions (Payer et al., 2003, Bortvin et al., 2004, Nakamura et al., 2007, Nakamura et al., 2012). STELLA was proposed to bind H3K9me2 found at imprinted regions, and protect them from TET-mediated demethylation (Nakamura et al., 2007, Nakamura et al., 2012, Bian and Yu, 2014). However, it has also been shown that STELLA can sequester UHRF1 in the cytoplasm, thereby preventing methylation by inhibiting the action of DNMT1 (Funaki et al., 2014, Mulholland et al., 2020, Li et al., 2018, Han et al., 2019). Therefore the roles of STELLA in ICR maintenance are complex and not yet entirely clear.

Maternally-methylated ICRs are found in promoter regions and are more prevalent, with the more than 20 maternal ICRs known in the mouse (Tucci et al., 2019). In contrast, there are three known paternally-methylated ICRs in the mouse, located in intergenic regions (Tucci et al., 2019). The methylation pattern at ICRs are in keeping with the

genome-wide patterns of methylation established in the maternal and paternal gametes; oocytes carry methylation mostly at genic regions, while sperm are methylated mostly at intergenic sequences (Kobayashi et al., 2012). Maternally-methylated ICRs commonly function by preventing the expression of ncRNA, thereby permitting maternal-allele expression of nearby protein-coding genes, while at the same locus on the paternal chromosome, the ncRNA is expressed and represses the protein-coding genes (Edwards and Ferguson-Smith, 2007). Paternally-methylated ICRs, though not located at promoters, also result in the nearby protein-coding genes being paternally expressed (Edwards and Ferguson-Smith, 2007). For example, in the case of the *IGF2/H19* cluster, the ICR is bound by CTCF and functions as an insulator element. Absence of methylation at the CTCF binding site on the maternal allele abolishes *IGF2* expression in favour of the nearby ncRNA, *H19*. Conversely, methylation at the ICR on the paternal chromosome prevents CTCF binding, thereby facilitating expression of *IGF2*. Based on these observations, it has been suggested that the ancestral state at imprinted loci may have been methylated, and that imprints may evolve via the loss of methylation to create the ICR (Edwards and Ferguson-Smith, 2007).

1.4.2 **Non-canonical imprints**

In addition to the canonical imprinted genes, a further class of loci inherit differential methylation patterns from the gametes (Proudhon et al., 2012, Kobayashi et al., 2012, Smith et al., 2012). These loci were termed transient germline DMRs (gDMRs) due to their differential methylation only persisting during the early embryonic stages (Proudhon et al., 2012, Kobayashi et al., 2012, Smith et al., 2012). Transient gDMRS had a similar distribution to canonical imprints, with oocyte- methylated loci in intragenic regions, and sperm-methylated loci in intergenic regions (Proudhon et al., 2012, Kobayashi et al., 2012). Like canonical imprinted regions, maintenance of methylation at transient gDMRs was suggested to occur via the binding of ZFP57 (Proudhon et al., 2012).

A recently-discovered type of imprinting occurs independently of DNA methylation and is instead mediated by oocyte-deposited H3K27me3 (Inoue et al., 2017a). These oocyte-derived marks repress the maternal allele, resulting in monoallelic gene expression from

the paternally-inherited chromosome in the pre-implantation embryo up until the blastocyst stage (Inoue et al., 2017a). Non-canonical imprinted expression was not retained in the post-implantation embryo, but some were retained in the extra-embryonic lineage (Inoue et al., 2017a, Hanna et al., 2019). It was subsequently discovered that H3K27me3 non-canonical imprints were localised at the long terminal repeats (LTRs) of endogenous retrovirus-K (ERV-K) insertions (Hanna et al., 2019). Differential marking of the parental alleles by H3K27me3 was not maintained in the post-implantation embryo (Inoue et al., 2017a, Hanna et al., 2019). However, in the extra-embryonic lineage, regions previously marked by allele-specific H3K27me3 instead acquired allele-specific DNA methylation (Hanna et al., 2019).

1.4.3 **Imprinting in metatherian mammals**

In keeping with the parental conflict theory for the evolution of imprinting, metatherian mammals, which like eutherians provide maternal resources to the developing embryo via the placenta, also display genomic imprinting. The search for imprinted loci in metatherians has so far involved the examination of loci known to be imprinted in eutherians for evidence of parent-of-origin dependent expression in metatherian species. This approach has identified eight imprinted loci in metatherians, and further confirmed that 13 loci imprinted in eutherians are biallelically expressed in metatherians (Stringer et al., 2014). In all instances where a metatherian imprinted gene has been identified, the parental expression 'direction' of the imprint has been conserved with that seen in eutherians. However, for some imprinted loci, expression was found to be strongly skewed towards one parental allele, rather than strictly monoallelic. For example, the *PEG1* gene (paternally expressed 1, also known as mesoderm specific transcript, *MEST*) was studied in the tammar wallaby *Macropus eugenii*, and the *PEG1α* isoform displayed paternally skewed biallelic expression in both somatic tissues and the placenta (Suzuki et al., 2005).

The mechanisms regulating metatherian imprinting are incompletely characterised. It was observed that a zinc finger protein with high homology to ZNF445 is present in the metatherian genome, and it was therefore speculated that this protein may have evolved together with genomic imprinting in therian mammals (Takahashi et al., 2019).

DNMT1O, the DNMT1 isoform produced in the oocyte and shown to be necessary for maintenance of imprints during one cell cycle of the early eutherian embryo, is also conserved in metatherians (Ding et al., 2003). However, ZFP57 and STELLA, two other eutherian proteins with links to imprinting regulation, are not conserved in metatherians (Takahashi et al., 2019). Given its essential role in eutherians, differential DNA methylation was an inviting candidate for a potential regulatory mechanism in metatherians.

Indeed, studies of some metatherian imprinted loci have uncovered evidence for the involvement of DNA methylation. For example, the *PEG10* (paternally expressed 10) gene was found to be monoallelically expressed from the paternal allele in the somatic tissue and placenta of the tammar wallaby (Suzuki et al., 2007b). The genomic arrangement of the *PEG10* locus was conserved between eutherians and metatherians, suggesting that the evolution of *PEG10* from a retrotransposon insertion event occurred in the common therian ancestor. A putative DMR was identified near the *PEG10* TSS; this site was methylated on the maternal allele, in line with paternal expression from this locus (Suzuki et al., 2007b). However, it is unknown if this site represents a true germline DMR, or is evident only in somatic cells (Suzuki et al., 2007b). There is one known instance of a germline DMR regulating metatherian imprinting. The *IGF2* (insulin-like growth factor 2) gene was found to be imprinted in two metatherian species, the South American opossum *Monodelphis domestica* and the North American opossum *Didelphis virginiana* (O'Neill et al., 2000, Suzuki et al., 2005). In both species *IGF2* was expressed from the paternal allele (O'Neill et al., 2000, Suzuki et al., 2005). The regulation of this locus was demonstrated to be conserved between eutherians and metatherians; in the tammar wallaby, homology search approaches identified a homologue of the ncRNA *H19*, and a germline DMR was found at CTCF sites located between *IGF2* and *H19* (Smits et al., 2008, Suzuki et al., 2013).

However, in the case of the *IGF2R* (insulin-like growth factor 2 receptor) gene, while the imprinted nature of the gene was conserved between metatherians and eutherians, the mechanism of regulation was not (Killian et al., 2000, Weidman et al., 2006a).
In other instances, genes known to be imprinted in eutherians have been found to display biallelic expression in metatherians. For example, *DLK1*, which is imprinted in

eutherians, was found to express biallelically in metatherians (Weidman et al., 2006b). Missing from the metatherian locus was the eutherian imprinted gene *MEG3*, suggesting that the arrival of *MEG3* via a LINE-1 integration event unique to eutherians launched the evolution of the *DLK1* locus as an imprinting domain (Weidman et al., 2006b). In a similar fashion, while *PEG10* is imprinted in metatherians, the neighbouring eutherian-imprinted genes *SGCE*, *PPPLR9A*, and *ASB4* are expressed biallelically (Suzuki et al., 2007b).

Together, these findings show that some imprinted loci are conserved between metatherians and eutherians, suggesting that the evolution of imprinting precedes the metatherian – eutherian divergence. However, only a subset of eutherian imprinted genes are conserved in metatherians, therefore additional imprinted genes evolved in the eutherian lineage. Regulation of imprinted genes in the eutherian lineage involves many complex imprinted domains. From what is known so far, it seems that metatherians have not evolved complex clusters of imprinted genes; with the exception of *IGF2/H19*, metatherian imprinted loci are all solo genes. However, it is clear that some aspects of the epigenetic regulation of imprinting evolved in the therian ancestor. Further study could identify a greater number of imprinted genes, and complex imprinting regulation in the metatherians. It is interesting to consider what other aspects of eutherian epigenetic regulation may be shared with metatherians.

1.5 Genome-wide reprogramming of DNA methylation

In contrast to the soma, where DNA methylation patterns remain relatively static, extensive changes to DNA methylation occur on two occasions during the early development of eutherian mammals. Both events involve global loss and subsequent re-establishment of DNA methylation and are referred to as genome-wide reprogramming (GWR). The first GWR event occurs in the cells of the germline, and the second occurs in the zygote and early preimplantation embryo. The nature of these GWR events has been the subject of much study, largely in the mouse.

1.5.1 **Genome-wide reprogramming in the mouse germline**

In the mouse, germ cells are specified in the epiblast, and then proliferate and migrate to the genital ridge by E10.5 (Ginsburg et al., 1990, Lawson and Hage, 1994, Ohinata et al., 2005). An early study making use of the differential capacity of the isoschizomers HpaII and MspI to digest methylated DNA revealed that mouse germ cells were globally less methylated than somatic cells, providing the first hint of GWR at this stage of development (Monk et al., 1987). The same study also demonstrated that the sperm had a higher global methylation level than the oocyte, implying that remethylation of germline cells occurs following GWR and establishes differing methylation patterns in male and female gametes (Monk et al., 1987). Supporting the notion that the gametes harboured differing methylation levels was the finding that some repeat sequences were lowly methylated in oocytes compared to sperm (Sanford et al., 1987, Walsh et al., 1998). Studies using methyl-sensitive restriction digestion followed by PCR to examine methylation levels of single loci confirmed that methylation was absent in germ cells at E12.5-13.5 and re-established later in development with different patterns in sperm and oocyte (Kafri et al., 1992). Regions associated with imprinted genes were also shown to carry low methylation in germ cells followed by remethylation during gametogenesis, thereby explaining how imprints could be erased and reset in a sex-specific manner with each generation (Chaillet et al., 1991, Kono et al., 1996, Tada et al., 1998, Ueda et al., 2000). The observation that genes involved in germ cell development were regulated by DNA methylation suggested that in addition to reprogramming of imprints, demethylation in the germline may function to facilitate transcription of genes necessary to gametogenesis and meiosis, which are otherwise repressed in somatic cells (Maatouk et al., 2006). However, it was not clear whether germ cells harboured high levels of DNA methylation and were subsequently demethylated, or if they were somehow set aside at specification, thereby avoiding somatic-like levels of methylation (Monk et al., 1987).

Bisulfite-sequencing of loci including imprinted and non-imprinted genes and repeat sequences established that germ cells initially carried high levels of methylation, and suggested that this was rapidly and actively erased upon arrival at the gonad, a process that was postulated to rely on DNA base-excision repair (BER; Hajkova et al., 2002, Hajkova et al., 2008, Hajkova et al., 2010). However, other studies suggested that

demethylation was evident in migrating germ cells prior to arrival in the genital ridges (Yamazaki et al., 2003, Seki et al., 2005). A study employing genome-wide bisulfite sequencing confirmed that germ cells isolated from the gonad were globally demethylated (Popp et al., 2010). The same study implicated the cytidine deaminase AID in a process of active demethylation, finding that AID-deficient germ cells retained higher levels of methylation than wild-type germ cells (Popp et al., 2010). However, significant demethylation was still observed in the absence of AID, implying that demethylation must occur by multiple mechanisms.

Subsequently, genome-wide studies examined further timepoints of germ cell development, and clarified that demethylation of PGCs occurred in two phases, the first a global passive loss of methylation by E9.5 (Guibert et al., 2012, Seisenberger et al., 2012), followed by TET-mediated active demethylation between E9.5-E13.5 that shapes the epigenome at targeted loci (Yamaguchi et al., 2012, Vincent et al., 2013, Hackett et al., 2012). These findings were in agreement with a study that suggested loss of methylation through passive dilution could account for germline reprogramming, based on the proliferation rate of germ cells and the repression of maintenance and *de novo* methylation machinery (Kagiwada et al., 2013). Therefore the picture that emerged was of germline GWR via passive dilution of methylation during migration, followed by some targeted active methylation, achieving an essentially unmethylated genome except for the retention of methylation at intracisternal A-particles (IAPs) and some LTR-ERV1 retroelements (Guibert et al., 2012). Recently however, it has been suggested that the role of TET proteins in germ cells may be to protect unmethylated sites against ectopic *de novo* methylation, rather than as part of the removal of methylation *per se* (Hill et al., 2018).

Genome-wide methods also allowed detailed study of the methylation patterns established in male and female gametes by the action of *de novo* methylation. Re-establishment of methylation was found to occur by E16.5 in the male germline and post-natally in the growing oocyte in the female germline (Kobayashi et al., 2012, Seisenberger et al., 2012, Molaro et al., 2014). *De novo* methylation of the germ cells relies on DNMT3A and DNMT3L (Bourc'his et al., 2001, Sakai et al., 2004, Bourc'his and Bestor,

2004, Kaneda et al., 2004, La Salle and Trasler, 2006, Vlachogiannis et al., 2015). In addition, the rodent-specific DNMT3C is involved in establishing methylation at evolutionarily young retrotransposons in the male germline (Barau et al., 2016). Correct establishment of the oocyte methylome also relies on the cytoplasmic retention of DNMT1, which occurs via the cytoplasmic sequestration of UHRF1 by STELLA, thereby preventing spurious establishment of methylation by DNMT1 (Li et al., 2018).

The methylation landscapes that result from *de novo* methylation of the germ cells differ dramatically, with the sperm globally hypermethylated relative to the oocyte. Distribution of methylation across the genome differs between the two germlines, with sperm methylation distributed throughout the genome with the exception of CpG-rich regions (Kobayashi et al., 2012). In contrast, methylation in the oocyte mostly occurs in gene bodies in a process linked to the oocyte transcriptional programme, including retroelement driven transcripts (Kobayashi et al., 2012, Chotalia et al., 2009, Smallwood et al., 2011, Veselovska et al., 2015, Brind'Amour et al., 2018). Germline DMRs (gDMRs) associated with imprinting control regions are established during the remethylation of the germline genomes. Oocytes bear a greater number of gDMRs than sperm, and the distribution of gDMRs occurs in manner consistent with the general methylation patterns of each gamete.

1.5.2 **Genome-wide reprogramming in the mouse zygote and preimplantation embryo**

As a consequence of the different programmes of *de novo* methylation during gametogenesis, the methylation states borne by the paternal and maternal genomes at fertilisation are vastly different. Immediately following fertilisation, and during preimplantation development, dramatic changes in methylation patterns are observed.

An early study comparing the differential patterns generated by HpaII and MspI digestion of genomic DNA suggested that the sperm was more highly methylated than the oocyte or the blastocyst, implying that methylation was lost during preimplantation development (Monk et al., 1987). The same study demonstrated that by E7.5, methylation was

regained, though the extra-embryonic compartment did not attain levels as high as the embryonic region (Monk et al., 1987). Other studies examined the methylation level of repetitive regions and found them to undergo demethylation in the early embryo, although with different dynamics for different repeat classes, and avoidance of demethylation by IAPs (Sanford et al., 1987, Howlett and Reik, 1991, Lane et al., 2003). Examination of several single-copy loci by methylation-sensitive restriction digestion and PCR confirmed that the methylation patterns inherited from the gametes were erased, and established that this occurred rapidly; by the 16-cell stage all tested CpG sites had undergone demethylation, with many sites already demethylated in 8-cell embryos (Kafri et al., 1992). Re-establishment of methylation was observed in the E6.5 embryo around the time of gastrulation (Kafri et al., 1992).

Methylation dynamics of the paternal and maternal genomes in the developmental stages immediately following fertilisation were revealed by immunostainings for methylated cytosine (Rougier et al., 1998, Mayer et al., 2000, Santos et al., 2002). These experiments revealed that the paternal genome rapidly lost methylation in the zygote, and that this preceded the time of DNA replication, implying that methylation was actively removed (Mayer et al., 2000, Santos et al., 2002). From the 2-cell stage onwards, methylation staining was also lost from the maternal chromosomes (Rougier et al., 1998, Santos et al., 2002). The presence of asymmetrically stained chromatids during this period was interpreted to represent replication-based dilution of methylation (Rougier et al., 1998). Bisulfite PCR experiments on several loci in mouse zygotes also indicated demethylation of the paternal allele prior to DNA replication, supporting the idea that methylation was actively removed from the paternal pronucleus (Oswald et al., 2000).

Reduced representation bisulfite sequencing of mouse gametes, zygotes, and embryos confirmed in base-specific, genome-wide detail the methylation dynamics shown by previous studies (Smith et al., 2012). The paternal genome was found to dramatically lose methylation in the transition from gamete to zygote, after which embryos exhibited a sustained phase of low methylation resembling the oocyte (Smith et al., 2012). The activation of the embryonic genome at the two-cell stage was therefore achieved in a genomic environment of particularly low methylation. Further decreases in methylation

were observed from the cleavage stages to the inner cell mass (ICM) of the blastocyst, which harboured very low methylation (Smith et al., 2012). Various classes of genomic repeats, including LINEs, SINEs and some LTRs were observed to lose methylation during the hypomethylated phase, although IAPs retained methylation as previously reported (Smith et al., 2012). A further finding of this study was the preponderance of novel DMRs conferred by the gametes and maintained during early development (Smith et al., 2012). Oocyte-contributed DMRs were enriched for CpG islands, while sperm-contributed DMRs featured mostly intergenic regions, and any possible functions for these DMRs were unknown (Smith et al., 2012).

Several mechanisms were suggested to facilitate active demethylation of mammalian genomes (Ooi and Bestor, 2008, Tahiliani et al., 2009, Ito et al., 2010, He et al., 2011, Ito et al., 2011) but it was not clear if any contributed to demethylation of the paternal genome in zygotes. The observation that components of the base-excision repair pathway were enriched in the paternal pronucleus of the zygote suggested that replacement of methylated cytosines with unmethylated cytosines could facilitate active paternal genome demethylation (Hajkova et al., 2010, Wossidlo et al., 2010, Santos et al., 2013). Knockdown of the elongator complex component Elp3 was demonstrated to impair demethylation of the paternal pronucleus in a live-monitoring assay for methylation levels in cultured zygotes (Okada et al., 2010).

The discovery that TET enzymes can induce demethylation prompted the investigation of TET-mediated oxidation as a potential mechanism for active demethylation in the zygote (Tahiliani et al., 2009, Ito et al., 2010, He et al., 2011, Ito et al., 2011). Several studies examined the distribution of 5hmC in zygotes by immunostaining, and observed an enrichment on the male pronucleus concomitant with the decrease of 5mC and prior to DNA replication, indicating that oxidation to 5hmC could account for the active loss of 5mC from the paternal pronucleus (Gu et al., 2011b, Iqbal et al., 2011, Wossidlo et al., 2011). 5fC and 5caC were also found be enriched on the paternal pronucleus on zygotes at the same time as 5mC signal was lost (Inoue et al., 2011). Expression profiling of *Tet* genes suggested *Tet3* could be responsible for hydroxymethylation in the zygote, and indeed depletion of TET3 abolished 5hmC on the paternal pronucleus (Gu et al., 2011b).

The maternal pronucleus was suggested to be protected from 5hmC conversion via the binding of STELLA to H3K9me2 (Nakamura et al., 2012).

These findings led to a model wherein the paternal pronucleus was actively demethylated by the action of TET3, followed by further passive dilution of methylation from both genomes thanks to the exclusion of the maintenance methyltransferase DNMT1 from the nucleus (Howell et al., 2001). However, 5hmC, 5fC, and 5caC persisted in stages after the zygote, decreasing in a manner consistent with passive dilution rather than active replacement with unmodified cytosine (Iqbal et al., 2011, Inoue et al., 2011). Other studies observed 5hmC to increase only at the time of DNA replication, temporally uncoupled from the prior decrease in 5mC, suggesting that other active demethylation mechanisms must be acting in addition to TET3 (Salvaing et al., 2012, Santos et al., 2013).

Subsequently, several high-resolution bisulfite sequencing studies examined the role of TET3 in the zygote. These studies found that the maternal genome was subject to a limited amount of active demethylation by TET3 in addition to the paternal genome, but that the contribution of TET3 to global demethylation was minor. Therefore it was concluded that other, possibly functionally redundant mechanisms must be at play, including passive demethylation and undefined active mechanisms (Peat et al., 2014, Wang et al., 2014a, Guo et al., 2014a, Shen et al., 2014). TET3 was implicated in targeted demethylation of certain loci, which fit with findings from knockout studies that showed TET3 loss did not prevent pre-implantation development (Tsukada et al., 2015, Inoue et al., 2015). Recently, it was suggested that the function of TET3 in the early embryo was to ensure the hypomethylation of certain loci by preventing the accumulation of *de novo* methylation at these sites (Amouroux et al., 2016).

Following the hypomethylated pre-implantation phase, a major remethylation occurs in the transition from blastocyst to post-implantation embryo. This occurs rapidly, with the E6.5 epiblast reaching adult-like methylation levels with a canonical bimodal distribution (Auclair et al., 2014, Smith et al., 2012, Zhang et al., 2018). Remethylation relies on the *de novo* methyltransferases DNMT3A and DNMT3B in a largely redundant manner, although differences were observed between the enzymes in their preferential targeting

of genomic features (Borgel et al., 2010, Auclair et al., 2014). Promoters of genes expressed only in the germline become methylated, along with some developmental genes and genes on the silent X chromosome in females (Auclair et al., 2014). During this phase, gDMRS present in the pre-implantation embryo are resolved, with the exception of DMRs associated with imprinted loci, which are protected from the *de novo* methylation wave to sustain the imprints into adulthood (Borgel et al., 2010, Proudhon et al., 2012, Smith et al., 2012). The cells of the E5.5 epiblast are pluripotent, therefore the methylation pattern established at this stage is propagated to all the somatic lineages. While the methylation patterns of somatic tissues are largely similar, differences are observed, largely at *cis*-regulatory elements, implying that local remodelling of methylation does occur associated with somatic differentiation (Hon et al., 2013).

The extra-embryonic lineage, though increasing in methylation level compared to the pre-implantation blastocyst, remains globally hypomethylated, with most regions attaining intermediate methylation levels rather than the canonical hypermethylation seen in adult tissues (Smith et al., 2017, Zhang et al., 2018). This hypomethylation is associated with lower expression of the *de novo* DNA methyltransferases (Zhang et al., 2018). Those sites that do attain high levels of methylation in extra-embryonic cells were observed to occur at CGIs near transcriptional start sites, and large regions near developmental genes that are typically found to be unmethylated in other tissues (methylation canyons; Smith et al., 2017, Zhang et al., 2018). Differences in methylation between the extra-embryonic and embryonic lineages might therefore represent reciprocal repression of developmental genes required in opposing lineages (Zhang et al., 2018). Non-canonical imprints carried by H3K27me3 from the oocyte become converted to DMRs in the extra-embryonic lineage during the remethylation phase (Hanna et al., 2019).

In addition to the action of the *de novo* methyltransferases, TET proteins are active during early development. While *Tet3* is expressed immediately following fertilisation, *Tet1* and *Tet2* become the predominant TETs expressed in the blastocyst. Single knockout embryos for each TET enzyme are able to develop to term. However, *Tet3* knockout embryos exhibit perinatal lethality, and *Tet1* and *Tet2* knockouts display later defects in neural progenitor cells and myeloid cells respectively (Dawlaty et al., 2011, Zhang et al., 2013a,

Rudenko et al., 2013, Ko et al., 2011, Li et al., 2011, Moran-Crusio et al., 2011). Double knockout of *Tet1* and *Tet2* results in variably penetrant mid-gestation developmental abnormalities and perinatal lethality (Dawlaty et al., 2013). These embryos displayed aberrant gene expression, loss of hydroxymethylation and an increase in methylation. TET1 and TET2 were shown to be involved in establishing methylation canyons at developmental loci in the epiblast, and at promoters and putative enhancers in the different germ layers at gastrulation (Zhang et al., 2018).

Tet1/*Tet3* double knockout was shown to cause transcriptional dysregulation in the preimplantation embryo and embryonic lethality at a range of developmental stages up to mid-gestation (Kang et al., 2015). Triple knockout of *Tet 1/2/3* in ESCs prevented differentiation and resulted in promoter hypermethylation and gene expression dysregulation (Dawlaty et al., 2014). Therefore TET enzymes seem to function by shaping the methylome to facilitate lineage specification and differentiation, in a partially functionally redundant manner.

1.5.3 **Conservation of genome-wide reprogramming in other eutherian mammals**

While GWR has been most comprehensively investigated in the mouse, studies have also examined this phenomenon in diverse eutherian species. Immunostaining of zygotes and early embryos revealed that demethylation of the paternal pronucleus is conserved in rat, cow, and human embryos (Barton et al., 2001, Zaitseva et al., 2007, Dean et al., 2001, Beaujean et al., 2004, Lepikhov et al., 2008, Fulka et al., 2004). Demethylation was also observed in the embryos of pigs, rabbits, goats and sheep, although with disagreement between studies in some instances (Chen et al., 2004, Beaujean et al., 2004, Shi et al., 2004, Reis Silva et al., 2011, Zhang et al., 2005, Lepikhov et al., 2008, Fulka et al., 2006, Jeong et al., 2007, Deshmukh et al., 2011, Park et al., 2010, Hou et al., 2005, Hou et al., 2008, Masala et al., 2017). Differences in timing and extent of demethylation, or in experimental conditions, may have accounted for these discrepancies.

More recently, genome-wide bisulfite sequencing studies have confirmed that cow, monkey, and human embryos all undergo a GWR event soon after fertilisation (Jiang et al., 2018, Duan et al., 2019, Gao et al., 2017, Guo et al., 2014b, Smith et al., 2014, Okae

et al., 2014, Zhu et al., 2018). Methylation levels in the early development of the cow were similar to those described in mouse, with sperm having high methylation levels and oocytes and early embryos being hypomethylated (Jiang et al., 2018, Duan et al., 2019). Methylation levels had dropped by the two-cell stage but reached a low point at the eight-cell stage, concurrent with EGA in this species (Duan et al., 2019). Methylation levels were seen to begin to rise again by the 16-cell stage, but as this was the latest timepoint profiled, any further view on remethylation remained elusive (Duan et al., 2019). In *in vitro* fertilised (IVF) monkey embryos, both the paternal and maternal genomes underwent active demethylation in the zygote and reached a methylation low point in the two-cell embryo (Gao et al., 2017). By the eight-cell stage, *de novo* methylation occurring concomitantly became the predominant force, and methylation levels rose slightly, before falling again in the morula and inner cell mass (Gao et al., 2017). However, the authors noted that many IVF embryos failed to develop beyond the eight-cell stage, a block that was less apparent in *in vivo* embryos (Gao et al., 2017). The eight-cell block could be overcome by inhibiting DNMT3A and 3B using morpholino oligonucleotides, thereby suggesting that *de novo* methylation at the eight-cell stage may have been artefactual (Gao et al., 2017).

In humans, sperm and somatic cells harboured highly methylated genomes in a bimodal distribution, while oocytes and early embryos were broadly hypomethylated (Guo et al., 2014b). Major demethylation of the genome was observed by the two-cell stage, with the paternal genome demethylating faster than the maternal genome (Smith et al., 2014, Zhu et al., 2018). Human embryos continued to lose methylation until the four-cell stage, followed by a slight increase at the eight-cell stage, and subsequently reached a low point in the blastocyst (Zhu et al., 2018). *De novo* methylation was observed during the same period as global demethylation, largely targeting young repetitive elements that retained a higher-than-average methylation level during early development, perhaps to ensure their repression (Guo et al., 2014b, Smith et al., 2014, Zhu et al., 2018). Therefore, methylation dynamics in the early human embryo resulted from a fine-tuned balance of methylation establishment and removal (Zhu et al., 2018). Similar to observations in the mouse, the human embryo harboured a set of maternally contributed DMRs which were transiently maintained during pre-implantation development (Smith et al., 2014).

Transient DMRs from mouse and human embryos showed very little syntenic overlap, implying that whatever the role for these signatures may be, they are regulated in a species-specific manner (Smith et al., 2014). Three further studies examined methylation in human germ cells during embryogenesis, revealing that humans germ cells undergo a robust GWR event, reaching a nadir of methylation by seven weeks gestation (Gkountela et al., 2015, Tang et al., 2015, Guo et al., 2015a). Evolutionarily young repeat elements retained slightly higher methylation than the rest of the genome (Gkountela et al., 2015, Tang et al., 2015, Guo et al., 2015a). Therefore, GWR is a conserved feature of eutherian mammal early development, albeit with differences in timing.

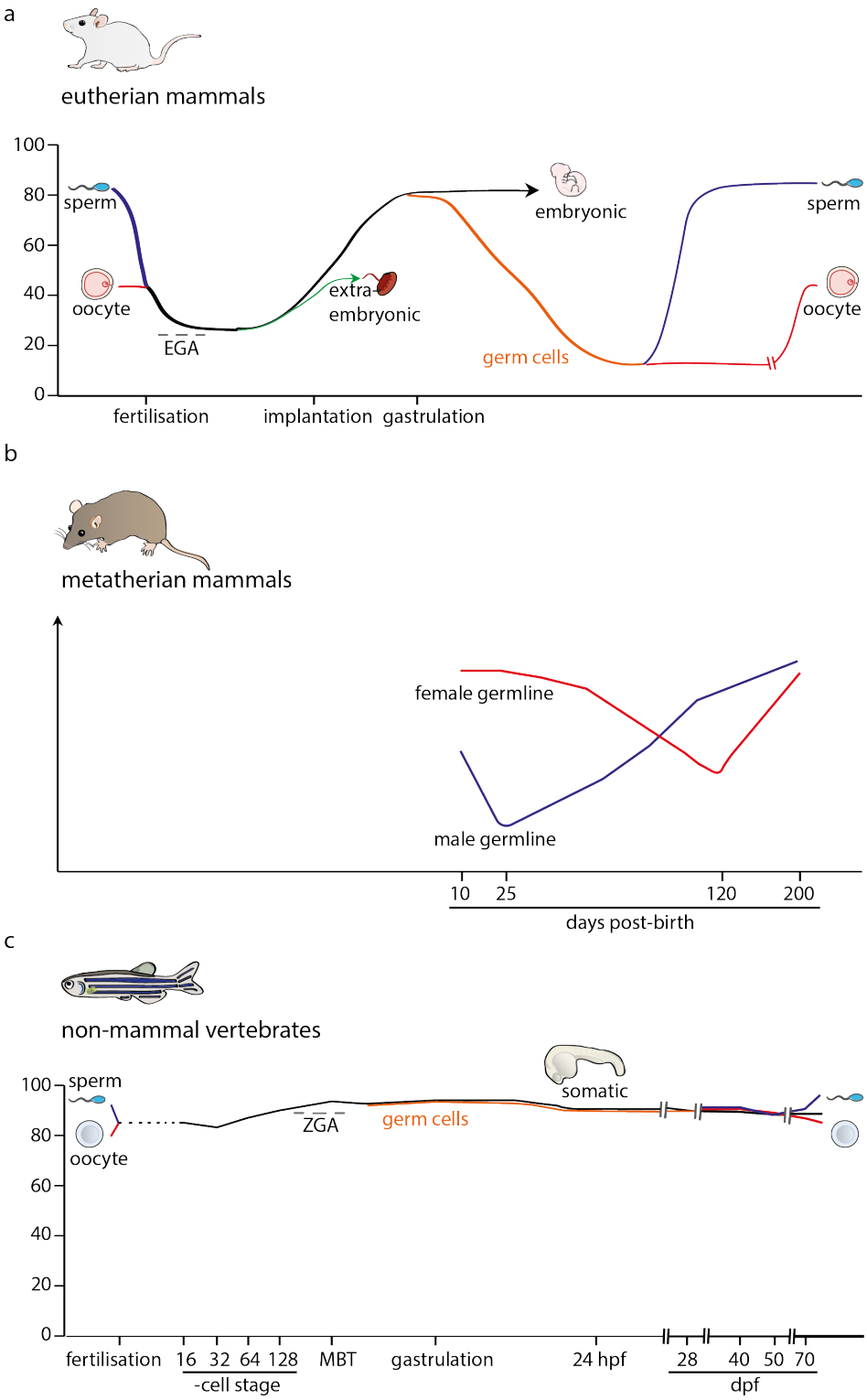


Figure 4: DNA methylation dynamics in mammalian and non-mammalian vertebrates. Representations of the methylation dynamics during early embryogenesis and germline development in different vertebrate lineages. **a**: Eutherian mammals, after (Smith et al., 2012, Seisenberger et al., 2012). **b**: Metatherian mammals, after (Ishihara et al., 2019). **c**: non-mammal vertebrates, after (Jiang et al., 2013, Skvortsova et al., 2019, Ortega-Recalde et al., 2019). dpf: hours post-fertilisation. dpf: days post-fertilisation.

1.5.4 **Methylation dynamics during development in non-mammalian vertebrates**

An early study using methylation-sensitive restriction digestion followed by Southern blot demonstrated that the *Xenopus laevis* genome is highly methylated throughout embryonic development (Veenstra and Wolffe, 2001). Another study examined methylation of *Xenopus* embryos by immunostaining and found that the genome does not undergo demethylation after fertilisation or during early cleavage stages. However, methylation was observed to decrease in embryos at the early blastula and mid-gastrula stages (Stancheva et al., 2002). Bisulfite sequencing demonstrated that the promoters of some single-copy genes were highly methylated in gametes and early embryos, and become demethylated at the gastrula stage, while others increased in methylation or remained the same (Stancheva et al., 2002). These findings suggested that demethylation represented local reprogramming of developmental promoters needed at the mid-blastula transition (Stancheva et al., 2002).

High levels of methylation were retained during early development of the zebrafish *Danio rerio*, based on methylation-sensitive restriction digestion plus Southern blots for a repeat element comprising 10% of the genome, with analysis of three single-copy genes supporting this finding (Macleod et al., 1999). Two landmark studies carried out whole-genome bisulfite sequencing of a time-course of zebrafish development, and revealed that unlike eutherian mammals, in zebrafish the highly-methylated paternal genome retains its methylation pattern throughout early development (Potok et al., 2013, Jiang et al., 2013). The hypomethylated oocyte genome was gradually reprogrammed, by a combination of locally targeted *de novo* methylation and passive demethylation, to reach a sperm-like pattern by the time of zygotic genome activation (ZGA) at the mid-blastula transition (Potok et al., 2013, Jiang et al., 2013). Reprogramming of the maternal genome was achieved even in parthenogenetic embryos, indicating that the sperm methylome was not required as a template (Potok et al., 2013). The sperm methylation pattern was very similar to that of somatic tissues, and the transition to this state by the maternal genome may therefore be necessary to attain a genome competent to undergo EGA and subsequent gastrulation (Potok et al., 2013). Supporting this idea, inhibition of DNA methyltransferases caused precocious transcription of genes involved in gastrulation

(Potok et al., 2013). Two studies investigated the germline of the zebrafish and found that DNA methylation was retained at levels as high as that found in the soma (Ortega-Recalde et al., 2019, Skvortsova et al., 2019). Therefore the prevailing view has been that GWR events are absent from the life-cycle of non-mammalian vertebrates.

In contrast to this view, examination of methylation in early embryos of the medaka *Oryzias latipes* using ELISA (Enzyme-Linked ImmunoSorbent Assay) found evidence of a decrease in methylation from sperm to zygote, concomitant with an increase in 5hmC (Wang and Bhandari, 2019a). WGBS of a time-course of germ cell development in the medaka suggested that global methylation levels in these cells was considerably lower (~40%) than in somatic cells (~80%), and that lowly-methylated stages coincided with an increase in expression of *tet*2 (Wang and Bhandari, 2019b). DNA methylation levels during early development of the mangrove killifish *Kryptolebias marmoratus* were examined using LUMA (LUminometric Methylation Assay), and found to decrease from ~ 30% in the zygote to a low of ~ 15% in the gastrula, before increasing to ~70% in the differentiated tissues (Fellous et al., 2018).

Methylation dynamics in the zebrafish germline and early embryo may therefore not be representative of all non-mammalian vertebrates. However, the studies in medaka, killifish, and frogs have been criticised based on their use of measures of global methylation (Ortega-Recalde and Hore, 2019). The embryos of fish and frogs are known to harbour very high levels of mtDNA and amplification of rDNA, both of which are typically found to be hypomethylated (Bird et al., 1981, Otten et al., 2016). This renders global measures of DNA methylation using ELISA and LUMA vulnerable to misinterpretation (Ortega-Recalde and Hore, 2019, Veenstra and Wolffe, 2001). Notably, the study that found retention of high methylation levels in frog embryos by Southern blot utilised careful controls for mtDNA (Veenstra and Wolffe, 2001). Whether methylation reprogramming strategies vary amongst the non-mammalian vertebrates remains to be clarified with further base-resolution studies.

1.5.5 Methylation dynamics during development in metatherian mammals

DNA methylation dynamics have been examined in the germline of a metatherian mammal, the tammar wallaby *Macropus eugenii*, but have not yet been studied in the metatherian early embryo. In the tammar wallaby, germ cells arrive at the genital ridge just before birth and continue to proliferate there until 25 days post-partum, when female cells begin meiosis and male cells enter mitotic arrest (Ullmann et al., 1997). Using bisulfite PCR, the methylation state of the *H19* imprinted region was examined in male tammar germ cells, clearly demonstrating that this locus was demethylated by 14 days after the arrival of germ cells in genital ridge (Suzuki et al., 2013). Re-establishment of methylation was observed by 34 days post-partum (Suzuki et al., 2013). The same study examined methylation of two repeat elements in the tammar genome by bisulfite PCR and suggested that some loss of methylation is observed over the same period (Suzuki et al., 2013). A subsequent study used anti-methylcytosine immunostaining to examine global methylation levels in the developing germline of male and female tammar wallabies, and found that the staining intensity of germ cells decreased relative to surrounding somatic cells between days 25-100 post-partum in males and days 25-200 post-partum in females (Ishihara et al., 2019). Therefore the current view holds that the germline of metatherian mammals undergoes a GWR event, but without further investigation using genome-wide, base-specific methods it is unknown to what degree and at which sequences. Meanwhile, it is entirely unknown if metatherian embryos undergo DNA methylation reprograming post-fertilisation.

1.5.6 Hypotheses to explain the evolution of genome-wide reprogramming in mammals

It is interesting to consider what function(s) may be served by global demethylation of both the germline and the early embryo. Based on current knowledge, it seems likely that GWR is a phenomenon specific to mammals, and therefore may have evolved to play a role related to mammal-specific aspects of development.

While it is clear from knockout studies that DNA methylation is required for successful development (Li et al., 1992b, Okano et al., 1999), it is less clear whether there is an

absolute requirement for global demethylation in the early embryo. So far there has been no report of the prevention of global loss of methylation in early development, probably because of functional redundancy in pathways that remove methylation. There is some suggestion that development can be successfully achieved without a fully demethylated paternal genome. In mouse embryos derived from round spermatid injection (ROSI), the paternal genome is demethylated after injection but gains methylation again during S-phase, implying that mature sperm carries the ability to resist methylation activity from the ooplasm (Kishigami et al., 2006, Polanski et al., 2008). Live offspring were successfully derived from ROSI embryos, indicating that high levels of methylation on the paternal genome were not incompatible with development (Kishigami et al., 2006, Polanski et al., 2008).

Several hypotheses for the role of GWR have been put forward; these are not necessarily mutually exclusive. The hypotheses relate to *1)* regulation of imprinted genes, *2)* preventing the transmission of epimutations, and *3)* establishment of transcriptionally permissive states during drastic cellular reprogramming during *a)* germline determination, and *b)* acquisition of totipotency in the zygote.

It is clear that one essential function of methylation loss in the germline is to reset the imprinted regions with parent-specific methylation. In eutherians mammals, knockout animals models in which the establishment of imprints is perturbed produce offspring with major developmental defects (Bourc'his et al., 2001). However, imprint resetting does not explain the global removal of methylation; the same result could be achieved by targeted removal at imprinted loci (Seisenberger et al., 2013). Indeed, the temporal decoupling of global germline demethylation by E9.5 and the later demethylation of imprinted loci suggests that different mechanisms may be responsible for these processes. The observation that genome-wide reprogramming also occurs in the germline of metatherian mammals, in which at least one imprinted gene possesses a gametic DMR, supports the notion that methylation must be reset in the germline to facilitate correct imprinting (Suzuki et al., 2013, Ishihara et al., 2019).

However, the resetting of imprints does not explain the second GWR event in the newly-fertilised embryo, as ICRs are protected from reprogramming at this stage. One possibility is that GWR in the early embryo represents the result of an evolutionary arms-race between the maternal and paternal genomes over imprinted genes (Reik and Walter, 2001a). In this scenario, the maternally-deposited oocyte cytoplasm acts as a defence system against paternally conferred imprints, rapidly stripping methylation from the sperm genome (Reik and Walter, 2001a). This idea is supported by the fact that most imprints are conferred by the oocyte, and by the protection of the maternal pronucleus from much of the active demethylation activity in the zygote (Reik and Walter, 2001a). Extant imprints are required for successful development and could therefore represent those sites where the arms-race reached equilibrium. Maternal defence against imprints could therefore have been the evolutionary stimulus for demethylation, and subsequently other selective pressures could have led to the co-option of GWR.

A second hypothesis suggests that global removal of methylation may serve to prevent the transmission of epimutations through generations (Seisenberger et al., 2013). Epimutations, i.e. acquired changes to the DNA methylation state, could be passed to the next generation if occurring in the germ cells. In mammals, epimutations could be acquired during two phases; first, during the period from fertilisation until germ cell induction; second, during the entire period of maturation, essentially up until gametes contribute to the next generation at fertilisation. This could explain the existence of the two GWR events in the mammalian lifecycle, one after germ cell specification, and one immediately following fertilisation. In support of this idea, there are few examples of transgenerational epigenetic inheritance in mammals (Morgan et al., 1999, Kazachenka et al., 2018). By comparison, in plants, where the germline is determined early and does not undergo a global demethylation, there are many examples of transgenerational epigenetic inheritance, and these can be inherited through many generations (Quadrana and Colot, 2016).

Finally, global reprogramming of the methylome may be associated with significant transitions in cell identity, during which the transcriptional program undergoes radical changes (Seisenberger et al., 2013). Such a transition occurs twice in early development;

at germline determination, and after fertilisation. The requirement for resetting of the epigenome in germ cells may be related to the mode in which the germline is determined. In mammals, the germline is induced from somatic cells of the epiblast via signalling. In fish, reptiles, and some amphibians, the germline is preformed; the germ cells are specified in the early embryo by maternally-inherited cytoplasmic factors. It has been suggested that in animals with an induced germline, a vast reprogramming of cell identity may be necessary to switch the selected cells from an epiblast to germ cell fate; reprogramming of the epigenome may be required to 'forget' the memory of the soma (Ortega-Recalde and Hore, 2019). By contrast, preformed germ cells do not transit through a somatic state and therefore may not require reprogramming, as observed in the zebrafish (Ortega-Recalde et al., 2019, Skvortsova et al., 2019). At fertilisation, the gametes represent terminally differentiated cell types, and therefore totipotency must be re-established in order to permit the development of all the cell lineages of the embryo (Hackett and Surani, 2013). At this stage, global loss of methylation may also facilitate the major transcriptional change that occurs with activation of the embryonic genome. However, global removal of DNA methylation cannot be an absolute requirement in the attainment of totipotency, or it would also be observed in non-mammalian organisms, where the gametes also represent terminally differentiated cell types.

The answer may lie in the relative developmental timing of the different organisms. Animals such as zebrafish, in which many cleavage divisions occur before EGA, are afforded the time to slowly reset their epigenomes to a transcriptionally competent state (Potok et al., 2013). In the mammalian embryo EGA occurs much earlier, after only a few cell divisions. The mammalian embryo is unique among animals in that it must also generate the extra-embryonic lineage; therefore the first cell lineage decision occurs earlier in mammals than other vertebrates. Early EGA in mammals is thought to be necessary to rapidly achieve totipotency ahead of the specification of the extra-embryonic lineage. Therefore, the rapid removal of methylation may create a permissive transcriptional environment in species without the time to slowly reprogram the epigenome (Potok et al., 2013, Surani, 2001, Hackett and Surani, 2013). Metatherian mammals also specify extra-embryonic lineages and activate their embryonic genome at

a relatively early stage. Examination of the methylome of the gametes and early embryos of metatherian mammals would shed light on these potential functions of GWR.

1.6 Mammalian XY sex chromosomes

During embryonic development, the gonads originate as bipotential structures with the ability to differentiate into a testis or an ovary. Sex determination via either an environmental or genetic system results in the subsequent development of either male or female characteristics (Capel, 2017). Genetic sex determination has the advantage of maintaining sex ratios in the face of environmental fluctuation. Mammals have an XY male/XX female genetic sex determination system, wherein presence of the Y chromosome results in testis formation. In prototherian mammals, there are five X and five Y chromosomes, without homology to the therian sex chromosomes (Grutzner et al., 2004, Rens et al., 2004, Veyrunes et al., 2008). In therian mammals, males are XY and females XX. In eutherians, the presence of the *Sry* gene on the Y chromosome results in the development of a testis (Koopman et al., 1991). In metatherians, presence of the Y chromosome also results in the development of a testis. An *SRY* homologue/orthologue is present, and is therefore likely to be the sex-determining factor (Harry et al., 1995). However, the role of metatherian *SRY* has never been formally tested by genetic studies.

1.6.1 **Evolution of the sex chromosomes**

The therian sex chromosomes evolved from an ancestral pair of autosomes prior to the metatherian-eutherian divergence (166-160 mya; Ohno, 1967, Charlesworth, 1996). The acquisition of *Sry* was likely the trigger of the divergence between the proto-X and proto-Y (Foster and Graves, 1994). Subsequently, selection to maintain *Sry* on the proto-Y resulted in suppression of recombination via chromosome inversions. In eutherians, the X and Y chromosomes harbour small homologous regions at which recombination is able to occur, ensuring correct segregation of the sex chromosomes at meiosis (Burgoyne, 1982). These are termed the pseudoautosomal region (PAR). In metatherians, there are no longer any recombining regions between the X and Y chromosomes, and a specialised cellular structure called the dense plate ensures their segregation at meiosis (Page et al.,

2006, Page et al., 2003). Most ancestral genes have therefore been lost from the Y chromosome, because the absence of recombination led to genetic drift (Lahn and Page, 1999).

In most cases where an ancestral gene has been retained on the Y chromosome, it exists as a degenerate version of its X-linked homologue (Skaletsky et al., 2003). The divergence between X and Y homologues has been used to identify four evolutionary strata resulting from chromosome inversion events (Lahn and Page, 1999). Ancestral Y-genes are often ubiquitously expressed and have roles in important cellular functions like transcription and translation; they are thought to have survived due to intolerance of alterations in gene dosage (Bellott et al., 2014, Cortez et al., 2014). In addition to the ancestral genes, the human Y chromosome also houses multiple ampliconic gene families, most of which are expressed in a testis-restricted pattern, in keeping with the prediction that as the Y chromosome is only ever present in males, Y-linked genes should evolve to hold roles in male fertility (Skaletsky et al., 2003). The human Y chromosome also harbours two blocks of sequence derived from the X chromosome via transposition events (Skaletsky et al., 2003). The mouse Y chromosome retains fewer ancestral genes than the human, but is largely composed of massively ampliconic acquired sequences (Soh et al., 2014). The metatherian Y chromosome has not been fully sequenced, but partial sequencing by two studies confirmed the retention of at least 18 ancestral genes (Bellott et al., 2014, Cortez et al., 2014).

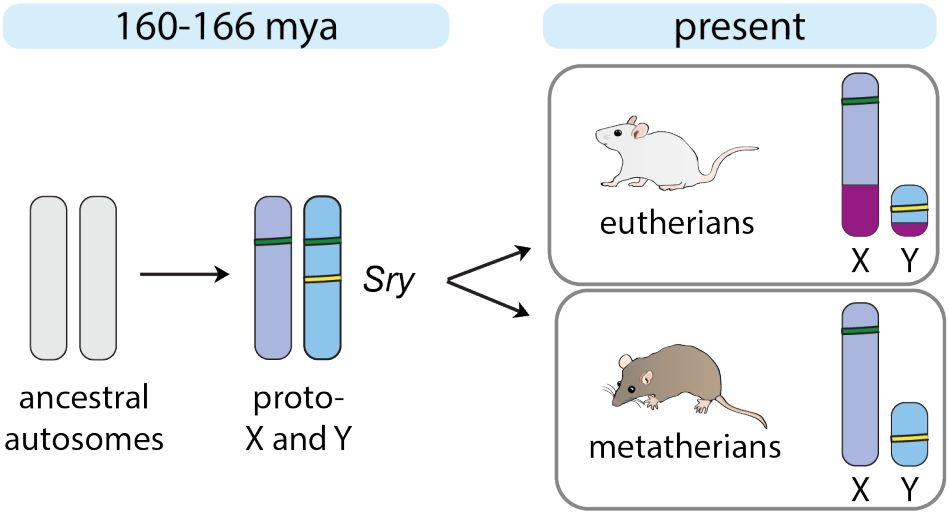


Figure 5: Evolution of the sex chromosomes of therian mammals. The therian sex chromosomes evolved from ancestral pair of autosomes following the acquisition of *Sry* (yellow). Suppression of recombination via chromosomal inversions lead to gene loss from the Y chromosome, while the X chromosome retained most ancestral genes (green) in both the eutherian and metatherian lineage (X conserved region, light purple). The X added region was acquired only in the eutherian lineage (dark purple).

The X chromosome retains a recombination partner during female meiosis, and therefore was not subject to gene loss. The majority of ancestral genes have therefore been retained on the X chromosome across therian species (Ross et al., 2005). In addition to these ancestral genes housed in the 'X-conserved region', a further acquisition after the metatherian-eutherian split lead to the 'X-added region' in eutherian lineage. The mouse X chromosome convergently acquired ampliconic sequence representing the same genes as the ampliconic Y-chromosome sequences; in both cases, these genes are exclusively expressed in the male germline (Soh et al., 2014).

1.7 X-chromosome dosage compensation

The loss of Y-linked genes resulted in dosage imbalance between the X chromosome and autosomes in males, and in X-linked gene products between males and females. Ohno (1967) hypothesised that in order to equalise the dose of X-linked genes relative to autosomes in males, the X chromosome would be expected to upregulate transcriptional output two-fold, thus reaching an X to autosome ratio (X:A) of 1. Ohno also postulated that X:A equalisation would result in two-fold overexpression of X-linked genes in the

female (Ohno, 1967). He hypothesised that female cells would thus have evolved a mechanism to repress X-linked gene expression and achieve dosage balance in comparison to males (Ohno, 1967).

1.7.1 **X-chromosome upregulation**

X chromosome upregulation (XCU), as predicted by Ohno, was initially reported to occur based on an analysis of a single locus in the mouse (Adler et al., 1997). Subsequently, analysis of human and mouse microarray data extended this finding to the entire X chromosome in many somatic tissues (Nguyen and Disteche, 2006). This finding was challenged by studies that did not observe an upregulation of X chromosome transcription in the mouse or the human (Xiong et al., 2010, Julien et al., 2012). However, these findings were confounded by the inclusion of genes that are highly expressed in reproductive tissues and lowly expressed elsewhere. These genes would not necessarily be subject to XCU, as they do not retain broad expression and dosage-sensitive functions. Examination of a restricted set of ubiquitously-expressed 'housekeeping' genes leant support to the occurrence of XCU (Pessia et al., 2012). The metatherian X chromosome also appear to display XCU (Julien et al., 2012).

1.7.2 **X-chromosome inactivation**

The inactivation of one X chromosome (XCI) in female mammals was identified by Lyon based on studies of the coat colour of mice (Lyon, 1961). In female mice heterozygous for coat colour mutations, patches of colour corresponding to both the wild-type and mutant allele were observed, and Lyon inferred that this resulted from the inactivation of one randomly-selected X chromosome in each cell at an early stage of development; the patches represented the clonal expansion of cells containing the same inactive X.

Subsequent studies have uncovered aspects of the regulation and developmental timing of XCI (Figure 6). Genetic studies of X-chromosome rearrangements and deletions had defined a minimal region required for XCI to occur in humans (X inactivation centre, *Xic*; Brown et al., 1991b). Subsequently, *Xist* was found to be localised in the *Xic* and to be transcribed exclusively from the inactive X (Brown et al., 1991a), and conservation of

Xist in the mouse was confirmed (Brockdorff et al., 1991). *Xist* was shown to be non-coding and localised in the nucleus, associating in *cis* with the inactive X chromosome (Clemson et al., 1996). The association of *Xist* RNA with the inactive X chromosome is not thought to result from direct contact with the DNA, as *Xist* remains associated with the nuclear matrix after removal of chromatin (Clemson et al., 1996). Deletion experiments subsequently proved that *Xist* was required for XCI to occur (Penny et al., 1996, Marahrens et al., 1997).

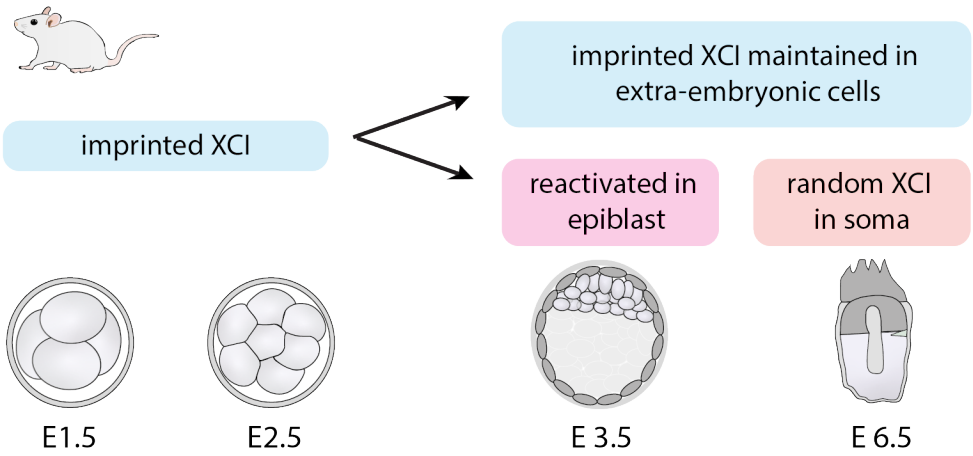
In the mouse, XCI occurs in two phases. From the four-cell stage until the blastocyst, XCI is imprinted, silencing the paternally-inherited X chromosome (Mak et al., 2004, Okamoto et al., 2004; Figure 6a). This chromosome is subsequently reactivated in the ICM of the blastocyst, while the imprinted inactivation is maintained in the extra-embryonic lineage (Takagi and Sasaki, 1975, West et al., 1977). As the embryo exits pluripotency, one X chromosome is chosen at random to be silenced in the cells of the epiblast (Takagi, 1974, Monk and Harper, 1978).

Imprinted paternal expression of *Xist* during imprinted XCI in the mouse was shown to dependent on a maternally-carried repressive mark (Tada et al., 2000), but to be independent of *de novo* DNA methylation in the oocyte (Chiba et al., 2008). *Tsix*, a ncRNA spanning the *Xist* locus and transcribed in antisense, was shown to be reciprocally expressed from the maternal X chromosome (Lee et al., 1999; Figure 6b). *Tsix* was subsequently shown to be required for imprinted XCI in the extra-embryonic tissues; *Tsix* mutations inherited from the mother are embryonic lethal (Lee, 2000, Sado et al., 2001). Absence of DNMT1O at the eight-cell stage caused hypomethylation of the *Tsix* regulatory region *Xite* on the paternal X chromosome, which resulted in biallelic expression of *Tsix* and failure of imprinted XCI in the extra-embryonic tissues (McGraw et al., 2013). However, *Tsix* is dispensable for correct imprinted expression of *Xist* in the preimplantation embryo (McGraw et al., 2013, Maclary et al., 2014). Therefore, DNA methylation imposed on *Xite* in the paternal germline seems to have a function specific to the extra-embryonic lineages, and is not sufficient to explain imprinted *Xist* expression. Recently, the nature of the *Xist* imprint was uncovered, with the discovery that H3K7me3

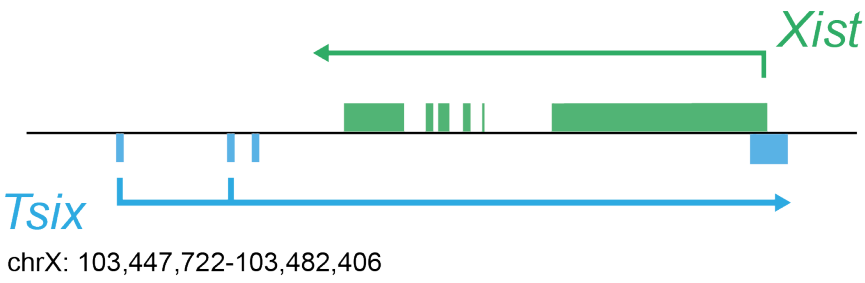
deposited at the *Xist* locus in oocytes ensures repression of the maternal allele during early development (Inoue et al., 2017b).

In other eutherian mammals, like the human and the rabbit, imprinted XCI does not occur, and random XCI takes place upon exit from pluripotency, as it does in the mouse (Okamoto et al., 2011). In human preimplantation embryos, *XIST* is expressed from the X chromosome in males, and from both X chromosomes in females, but does not lead to silencing of the chromosomes (Okamoto et al., 2011). Later, *XIST* is lost from the chromosome that will remain active, and remains on the chromosome chosen to be inactivated (Okamoto et al., 2011).

a


imprinted XCI
imprinted XCI maintained in extra-embryonic cells
reactivated in epiblast
random XCI in soma
E1.5
E2.5
E 3.5
E 6.5

b


Xist
Tsix
chrX: 103,447,722-103,482,406

c

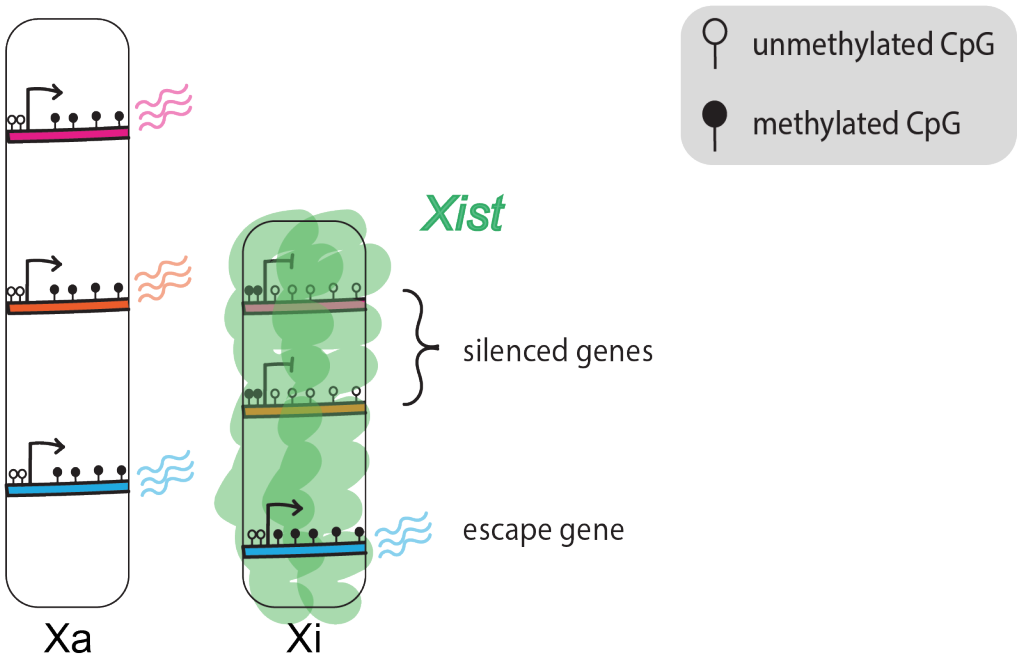

unmethylated CpG
methylated CpG
Xist
silenced genes
escape gene
Xa
Xi

Figure 6: Mouse X-chromosome inactivation.
Schematic representing aspect of X chromosome inactivation (XCI) in the mouse. **a**: Timing of XCI during early development. In the mouse, imprinted XCI occurs from the four-cell stage until the late blastocyst, where reactivation occurs in the cells of the epiblast. Imprinted XCI is maintained extra-embryonic lineages, while random XCI is established in the soma. **b**: Representation of the *Xist*/*Tsix* locus, indicating their overlapping antisense transcription. **c**: The inactive X is coated by *Xist* and repressive epigenetic modifications are recruited. Repressive DNA methylation is established at promoters of silent genes on the inactive X chromosome, but is missing from the promoters of escape genes.

Xist coating recruits a suite of repressive epigenetic marks to the inactive X chromosome. The inactive X is depleted of markers of active chromatin, like acetylated histones and trimethylation of H3K4 (Jeppesen and Turner, 1993, O'Neill et al., 2008). Conversely, the inactive X is enriched for repressive chromatin modifications, such as methylation of H3K9 and H3K27, as well as harbouring the repressive histone variant macroH2A (Boggs et al., 2002, Heard et al., 2001, Peters et al., 2002, Plath et al., 2003, Silva et al., 2003, Costanzi and Pehrson, 1998). Following the establishment of the repressive chromatin state, DNA methylation is laid down at CGI promoters of genes on the inactive X chromosome later in development, during the wave of remethylation associated with exit from pluripotency (Auclair et al., 2014; Figure 6c). In contrast to promoters, the gene bodies of silent genes on the inactive X are reported to be undermethylated (Hellman and Chess, 2007; Figure 6c). At a large proportion of these methylated sites, the acquisition of methylation is dependent on the structural maintenance of chromosomes protein SMCHD1 (Blewitt et al., 2008, Gendrel et al., 2012). The inactive X chromosome also adopts a unique 3D structure, largely devoid of topologically associated domains (TADs) (Giorgetti et al., 2016). It was recently shown that SMCHD1 is required for the inactive X to adopt this 3D conformation, via the exclusion of CTCF and cohesin (Gdula et al., 2019). In the absence of SMCHD1, TAD reformation was observed, alongside aberrant DNA methylation patterns and reactivation of genes on the silent X chromosome (Gdula et al., 2019).

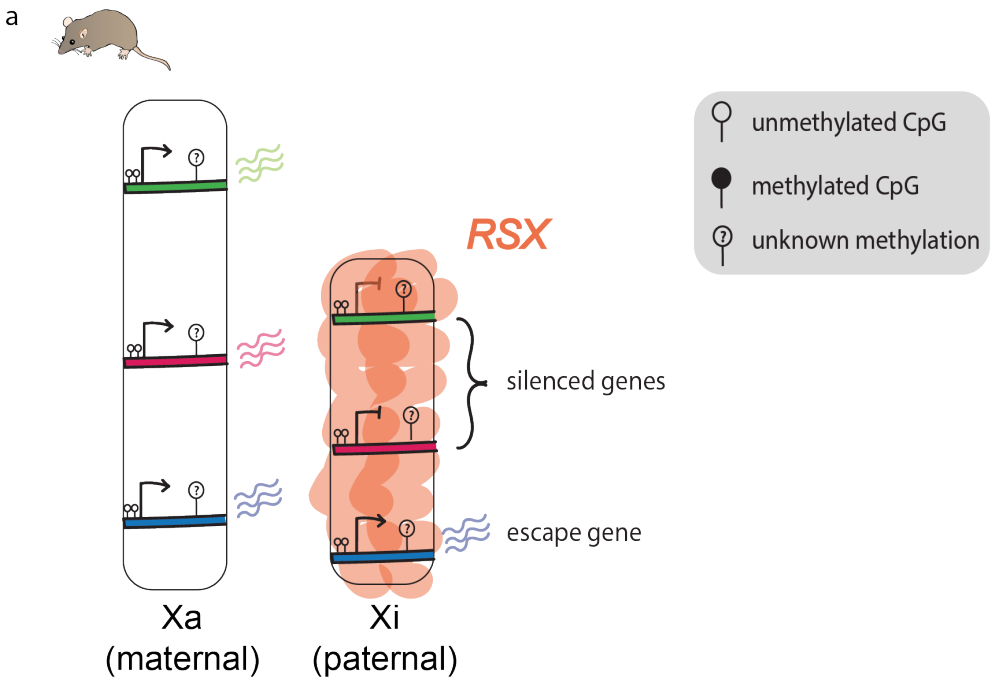
Not all X-linked genes are subject to XCI. 3-7% of mouse genes, and 8-15% of human genes escape from inactivation (Berletch et al., 2015, Carrel and Willard, 2005, Zhang et al., 2013b, Cotton et al., 2013). Escape genes avoid the repressive epigenetic states to which their silenced neighbours are subject. For example, escape genes are depleted for

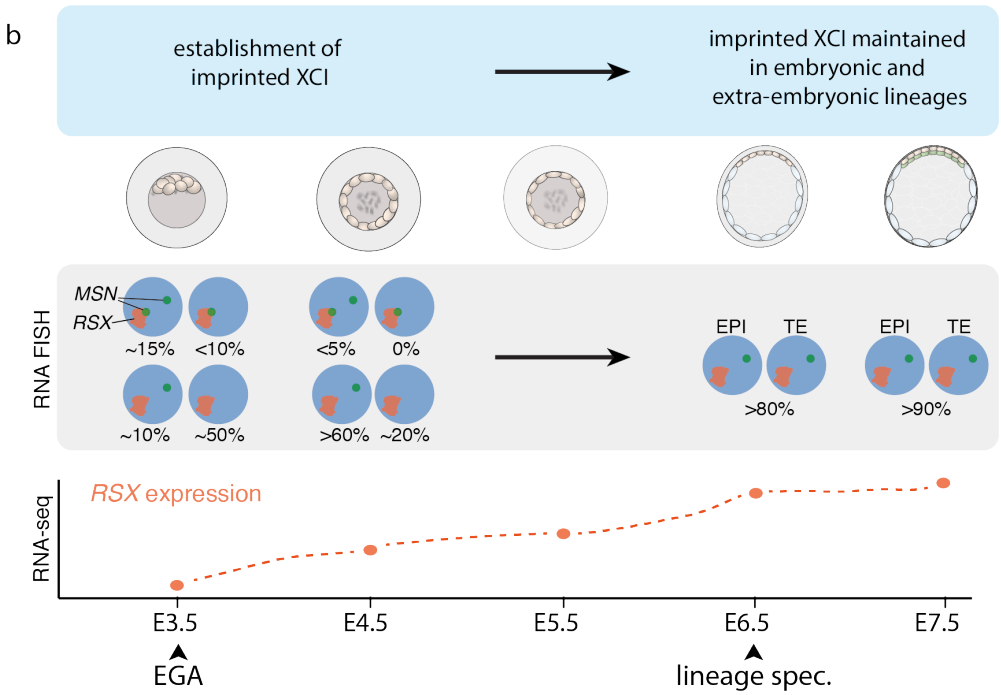
the repressive histone modification H3K27me3, and enriched for active marks such as histone acetylation and RNA polymerase occupancy. Escape genes reside towards the exterior of the 3D structure adopted by the inactive X chromosome and adopt a TAD structure, unlike the rest of the Xi (Chaumeil et al., 2006, Splinter et al., 2011, Giorgetti et al., 2016).

XCI also occurs in metatherian mammals, but unlike eutherians, is strictly imprinted, with the paternal X chromosome always silenced (Sharman, 1971; Figure 7). Unlike eutherian mammals, in metatharians there is no *Xist*. In fact, *Lnx3,* the gene that evolved to become *Xist* in the eutherian lineage retains protein-coding function in metatherians (Shevchenko et al., 2007, Hore et al., 2007, Duret et al., 2006). The discovery of *RSX*, a ncRNA with properties consistent with a role in XCI, suggested that both therian lineages have evolved convergent mechanisms of dosage compensation (Grant et al., 2012). *RSX* is expressed only in female cells, and derives exclusively from the paternal X chromosome, coating it in *cis* in a manner similar to *Xist* in eutherians (Grant et al., 2012; Figure 7a). Integration of *RSX* as a transgene onto an autosome in mouse embryonic stem cells resulted in *RSX* RNA spreading and silencing nearby genes, indicating that like *Xist*, *RSX* has the capability to induce silencing in *cis* (Grant et al., 2012).

Several studies have profiled aspects of the epigenetic landscape of the metatherian Xi, including DNA methylation and histone post-translational modification, and the findings suggest that some epigenetic features are shared by the metatherian and eutherian Xi, while others differ. In several studies using locus-specific bisulfite PCR, it was found that unlike eutherians, DNA methylation was absent from promoters on the metatherian Xi (Kaslow and Migeon, 1987, Loebel and Johnston, 1996, Wang et al., 2014b; Figure 7a). RBBS experiments recently extended this finding to a large proportion of X-linked promoters (Waters et al., 2018), however the methylation patterns at gene bodies and intergenic regions on the metatherian Xi remain mostly unexplored (Figure 7a). Several studies have examined the profile of various histone modifications on the metatherian X chromosome using immunofluorescence. These studies found that active histone marks including lysine acetylation and H3K4 methylation were depleted from the metatherian Xi, similar to eutherian mammals (Koina et al., 2009, Rens et al., 2010). Initially it was

suggested that the metatherian Xi was not enriched for repressive histone marks (Koina et al., 2009). However, subsequent studies established that the repressive marks H3K27me3, H3K9me3, and H4K20me3 were enriched on the metatherian Xi (Mahadevaiah et al., 2009, Rens et al., 2010). Similar to the eutherian Xi, some metatherian X-linked genes escape from XCI (Wang et al., 2014b). Using chromatin immunoprecipitation (ChIP) sequencing to investigate the profiles of H3K27me3 and H3K4me3 across the X chromosome, this study showed that XCI-escape genes harboured an active histone signature, in contrast to neighbouring silenced genes (Wang et al., 2014b). Therefore eutherian and metatherian XCI share similar repressive histone landscapes, but differ in regards to DNA methylation patterns (Figure 7a).


a
unmethylated CpG
methylated CpG
unknown methylation
RSX
silenced genes
escape gene
Xa
(maternal)
Xi
(paternal)


b
establishment of
imprinted XCI
imprinted XCI maintained
in embryonic and
extra-embryonic lineages
RNA FISH
MSN
RSX
~15%
<10%
<5%
0%
~10%
~50%
>60%
~20%
EPI
TE
>80%
EPI
TE
>90%
RNA-seq
RSX expression
E3.5
E4.5
E5.5
E6.5
E7.5
EGA
lineage spec.

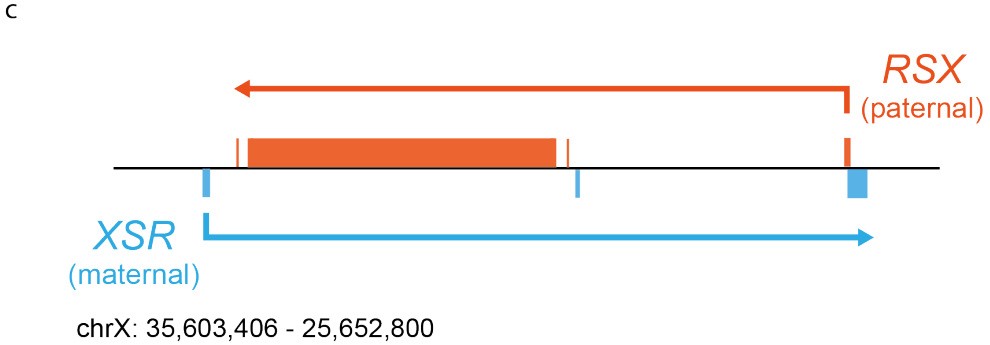

c
RSX
(paternal)
XSR
(maternal)
chrX: 35,603,406 - 25,652,800

Figure 7: Opossum X-chromosome inactivation.
Schematic representing aspect of X chromosome inactivation (XCI) in the opossum. **a**: Epigenetic landscape of the opossum X chromosome. The inactive X is coated by *RSX*. Repressive DNA methylation is not reported to occur at promoters of silent genes on the inactive X chromosome, and the methylation state of the wider chromosome is not described. **b**: Ontogeny of XCI during opossum early development. Imprinted XCI is established from the time of EGA at E3.5. *RSX* expression is apparent from E3.5, and is observed to form clouds by RNA FISH. By E4.5, monoallelic expression of the X-linked gene *MSN* is established in most cells. Imprinted XCI is maintained in both epiblast (EPI) and trophectoderm (TE) lineages. **c**: Representation of the *RSX*/*XSR* locus, indicating their overlapping antisense transcription units and parent-of-origin expression. Panels **b -c** are based on unpublished work by Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory.

A recent study examined the ontogeny of XCI in the early embryos of the South American opossum *M. domestica* (Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished). RNA-seq of a timecourse of early opossum development established that *RSX* was expressed at E3.5, at the same time as EGA (Figure 7b). Examination of *RSX* by RNA-FISH confirmed the timing of expression and revealed that *RSX* was exclusively expressed in female embryos and always originated from one chromosome (Figure 7b). X:A expression ratios were 1 at all the timepoints examined, indicating that both XCU and XCI are were rapidly established to ensure dosage compensation at the time of EGA. In support of this finding, RNA-FISH for the X-linked gene *MSN* showed that expression could be detected from the paternal X chromosome in E3.5 embryos, but was silenced by E4.5 (Figure 7b). In comparison to eutherians, where both X chromosomes are active in the epiblast, the opossum X chromosome remained silent following its inactivation at E3.5 – 4.5. *RSX* clouds were observed by RNA-FISH in epiblast and extra-embryonic cells of the E6.5 - 8.5 opossum embryo, and X:A autosome ratios remained stably at 1, indicating X inactivation was maintained throughout early development (Figure 7b).

While metatherian XCI is known to be imprinted, the mechanisms regulating this imprint are unknown. One candidate mechanism would be an epigenetic mark controlling the expression of *RSX,* established differentially between the parental germlines. RNA-sequencing of opossum embryos revealed the existence of an antisense ncRNA overlapping the *RSX* locus (Figure 7c). Termed *XSR*, this RNA was transcribed in early embryos, beginning at E3.5 and continuing to be expressed until at least E7.5, but was not expressed in adult tissues. *XSR* was expressed in both male and female embryos, and

by SNP-PCR analysis was shown to be expressed exclusively from the maternal X chromosome (Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished). *XSR* and *RSX* are therefore reciprocally imprinted ncRNAs (Figure 7c). Analogous to the *Xist*/*Tsix* paradigm in the mouse, *XSR* may function to prevent *RSX* expression from the maternal X chromosome, ensuring that it remains active. Meanwhile, the absence of *XSR* expression on the paternal X chromosome would permit *RSX* expression, leading to its silencing. An imprinted mechanism of XCI may have evolved to prevent biallelic X chromosome silencing in species in which XCI occurs very early. Such species might be particularly vulnerable to the negative effects of biallelic X-silencing due to the low cell number at the time of XCI. It is therefore of great interest to examine epigenetic regulatory marks at the *RSX*/*XSR* in the germ cells and early embryos of opossums, in order to uncover the molecular nature of the imprinting affecting this locus.

1.8 Aims of this thesis

Studying metatherians mammals can cast light on the evolution of important mammalian traits. In this thesis, I aimed to survey DNA methylation in the context of early development and X-chromosome inactivation of metatherians, and by comparison to eutherian species, to understand the evolution of roles for DNA methylation in the mammalian lineage. I aimed to assess *1)* the global methylation distribution of the metatherian X chromosome, and its relationship to gene expression, *2)* whether DNA methylation has a role in the regulation of imprinted XCI in metatherians and *3)* whether metatherians, like eutherians, dramatically reprogram the methylome in early embryogenesis.

To answer these questions, I have used bisulfite sequencing approaches to profile DNA methylation in the somatic tissues of adult male and female opossums, also generating datasets from equivalent mouse samples to facilitate comparison to eutherian mammals. Using a low-input bisulfite sequencing method, I have profiled DNA methylation in a developmental timecourse representing opossum gametes, early embryos, and blastocysts. The timecourse encompassed activation of the embryonic genome, onset of

X chromosome inactivation, and the specification of embryonic and extra-embryonic lineages in the opossum embryo, thereby shedding light on the role of DNA methylation in these key developmental events.

Chapter 2. Materials & Methods

2.1 Animal work

Mouse samples used in this project were taken from colonies maintained in the Francis Crick Institute Biological Research Facility (BRF). Opossum samples used for experiments presented in Chapter 3 and Chapter 5 were taken from the Francis Crick Institute BRF opossum colony. All animals from the Francis Crick Institute BRF were maintained according to the United Kingdom Animal Scientific Procedures Act (ASPA) 1986 and the ethics guidelines of the Francis Crick Institute. Opossum tissue samples used for the experiments presented in Chapter 4 were provided by John VandeBerg from an opossum colony maintained at the University of Texas under appropriate local regulations. Particular mouse and opossum crosses are outlined in the relevant chapters.

2.1.1 **Opossums**

2.1.1.1 General husbandry

Opossums maintained in the Francis Crick Institute BRF colony were individually housed in Double Decker cages (GR1800, Tecniplast), to allow climbing behaviour. With the exception of mating periods, adult male and female opossums were housed in separate rooms. Enrichment activities including tunnels, rodent balls, and nesting boxes were provided. Opossums had free access to water and dried food, and their diet was supplemented every second day by live mealworms, which also provided behavioural enrichment, and weekly by fresh fruit. The environment was maintained with temperature between 24-28°C, humidity between 55-75 %, and a light:dark cycle of 14 hours:10 hours. Animals were considered adults at six months.

2.1.1.2 Humane culling procedure

The opossum to be culled was anaesthetised by exposure to isoflurane gas until unresponsive to rocking of the cage and/or toe pinch. Subsequently, the animal was culled by cervical dislocation using a plastic scraper tool. Death was confirmed either by the severing of the carotid artery, or by observation of the onset of rigor mortis.

2.1.1.3 Matings

Matings were conducted in the male room. Adult males and females selected as breeding pairs were placed in single-storey rat cages immediately adjacent to each other for two days, then swapped into each other's cages for a further two days. Exposure to male pheromones via bedding in the male cage induced oestrous and rendered female opossums receptive to the paired males. Subsequently, males and females were placed into the same cage, and monitored for several hours for signs of aggression. If aggressive behaviour was displayed the pair were separated; otherwise, mating pairs were kept together for ten days. Following this period, animals were returned to their home cages, and females were monitored for the presence of a litter for 15 days.

2.1.1.4 Timed collection of oocytes and pre-implantation embryos

To facilitate timed collection of oocyte and embryo samples, animals were monitored by CCTV camera during the ten-day mating period. The occurrence of mating behaviours as defined by Baggot (1987) were taken to indicate pregnancy. To ensure consistency of developmental staging, embryos were only collected from matings that occurred between 21:00 and 24:00. Embryos were recovered from the uteri at E1.5 (36 hours post-coitum; hpc), E2.5 (60 hpc), E3.5 (84 hpc), E4.5 (108 hpc), E5.5 (132 hpc), E6.5 (156 hpc), E7.5 (180 hpc). In the case of oocyte collection, the female was mated to a vasectomised male, and oocytes were recovered from the uterus at 36 hpc. On the day of collection, the pregnant female was culled (see section 2.1.1.2.) between 10:00 and 12:00, and the uteri were removed into room temperature phosphate-buffered saline (PBS; Gibco #14190-094). The uteri were bisected using a scalpel blade, everted, and gently shaken under PBS to release pre-implantation embryos into the surrounding solution. Individual embryos were then located using a Leica MC80 dissecting microscope, aspirated using a Stripper pipette with a 600 µm tip (MXL3-STR and MXL3-600, Cooper Surgical), and washed three times through drops of clean PBS. Some E7.5 embryos were too large for a 600 µm pipette and were therefore handled with a conventional micropipette fitted with a 200 µL tip. Embryos were imaged using a Leica DMIL LED microscope with a 20x objective

and were then processed for different experiments as described in the relevant sections of this chapter (low-input BS-seq, section 2.9.1; immunostaining, section 2.11.2).

2.1.2 **Mice**

2.1.2.1 General husbandry

Mice maintained in the Francis Crick Institute BRF colony were housed in individually ventilated cages (GM500, Tecniplast) with automatic water, air, and humidity management systems, and a 12:12 light:dark cycle. Adult females and males were housed in separate cages in groups of 3-4 animals. Mice had free access to water and food, and enrichment activities including rodent balls and nesting boxes were provided.

2.1.2.2 Humane culling procedure

Mice were culled by cervical dislocation, and death was confirmed by the severing of the carotid artery, or by observation of the onset of rigor mortis.

2.1.2.3 Timed matings

Matings for the purpose of timed collection of embryos were conducted as follows: females were placed into the cage of males at approximately 17:00, and the following morning were examined for evidence of a vaginal plug. When a plug was observed, this was taken to indicate mating had occurred. 12:00 on the day a plug was observed was assumed to be E0.5.

2.1.2.4 Collection of pre-implantation blastocysts

For the collection of E3.5 pre-implantation mouse embryos, females were culled (section 2.1.2.2) at approximately 12:00 on the third day following the observation of a vaginal plug. The uteri were removed and rinsed under room temperature PBS (Gibco #14190-094). Using a 27G blunt-ended steel needle, embryos were flushed from the uteri into 0.5 ml PBS. Embryos were located under a Leica MC80 dissecting microscope, aspirated using a Stripper pipette with a 275 µm tip (MXL3-STR and MXL3-275, Cooper Surgical), and washed three times through drops of clean PBS. A Leica DMIL LED microscope with a 20x objective was used to image each embryo. Subsequently, embryos

were processed for different experiments as described in the relevant sections of this chapter (low-input RRBS, section 2.8.1.2; low-input BS-seq, section 2.9.1.2).

2.2 Expression analysis of DNA methylation genes

RNA-seq data generated for the analysis of allele-specific expression in opossum and mouse adult tissues (library preparation and initial data processing described in section 2.6) were used to assess the expression levels of genes involved in DNA methylation. Gene annotations were retrieved from Ensembl for mouse and opossum genomes (versions GRCm38.97 and monDom5.97 respectively). Reads overlapping genes were counted from bam files using the `featureCounts` function from the R package *Rsubread.* Multi-mapping reads were excluded from counting. Fragments Per Kilobase per Million mapped reads (FPKM) values were calculated using the `fpkm` function of the R package *DESeq2* using the robust median ratio method. $\log_2$(FPKM+1) values were calculated for genes of interest, and boxplots representing these values were generated for different tissues using the *ggplot2* package.

2.3 High-input reduced representation bisulfite sequencing

2.3.1 Extraction of gDNA from tissue samples

Following culling as detailed in section 2.1, mouse and opossum tissues were dissected, sliced into small pieces (~10 mm), and snap frozen in liquid nitrogen. Tissue pieces were then stored at -80°C until use. Frozen tissue was pulverised using a pestle and mortar pre-cooled on dry ice, and genomic DNA (gDNA) was extracted using the PureLink Genomic DNA Mini kit (#K18290-02, Invitrogen) following the manufacturer's instructions. Pulverised tissue was transferred to a 1.5 ml microcentrifuge tube with 180 µL PureLink digestion buffer and 20 µL Proteinase K and incubated at 55°C for 4 hours with occasional vortexing. Samples were centrifuged at 17,100 x *g* for 3 minutes to remove particulate materials, and the supernatant was transferred to a new microcentrifuge tube with 20 µL RNase A for 2 minutes at room temperature. 200 µL PureLink Genomic Lysis/Binding buffer was added and samples were briefly mixed by vortexing, followed by the addition of 200 µL 100% EtOH and a 5 second vortex and loading onto a PureLink Spin Column. Columns were centrifuged at 10,000 x *g* for 1 minute at room temperature,

then washed in 500 µL PureLink Wash Buffer 1 and centrifuged at 10,000 x *g* for 1 minutes at room temperature. The wash step was repeated with PureLink Wash Buffer 2. Purified gDNA was eluted into a clean microcentrifuge tube by 1-minute room temperature incubation in 30-100 µL PureLink Genomic Elution buffer followed by a 1-minute centrifugation at 17,100 x *g*. To confirm that gDNA was high molecular weight, 5 µL of each preparation was electrophoresed through a 2% (w/v) agarose Tris-acetate-EDTA (TAE) gel containing 0.5% (v/v) SYBR Safe (#SS33102, ThermoFisher) at 80V for 40-60 minutes, followed by visualisation of gDNA under UV light. The migration of gDNA preparations was compared to that of the GeneRuler DNA Ladder Mix (SM0331, ThermoFisher).

2.3.2 **Library preparation and sequencing**

RRBS libraries were prepared according to the method of Gu et al. (2011a) with several modifications. Approximately 1 µg of each gDNA preparation was digested for four hours at 37°C in 20 µL of 1x CutSmart buffer (#B7204S, NEB) containing MspI and BfaI restriction enzymes (10 units; # R0106S and #R0568S, NEB). To confirm that gDNA was successfully digested, 2 µL of each digestion was electrophoresed through a 2% (w/v) agarose TAE gel containing 0.5% (v/v) SYBR Safe (#SS33102, ThermoFisher) at 80V for 40-60 minutes, followed by visualisation of the digested DNA smear under UV light in comparison to the GeneRuler DNA Ladder Mix. The remaining digested DNA was purified as follows: the total sample volume was made up to 200 µL with nuclease-free water, an equal volume of phenol/chloroform/isoamyl alcohol was added, and the sample was mixed by gentle inversion. Samples were centrifuged for five minutes at 14,000 x *g*, and the aqueous phase was removed to a fresh tube. DNA was precipitated at -20°C for 30 minutes following the addition of GlycoBlue (#AM9515, ThermoFisher) to 50 ng/ml, NaCl to 0.2 M, and 2.5 volumes of 100% EtOH. Precipitated DNA was pelleted by centrifugation at 16,000 x *g* for 30 minutes at 4°C, the pellet was washed twice with 70% (v/v) EtOH, air dried at room temperature, and resuspended in 23 µL Elution Buffer (EB; #19086, Qiagen). DNA concentration was measured by Nanodrop fluorometer.

RRBS libraries were prepared from 300 ng of DNA using the NEBNext Ultra DNA Library Prep Kit for Illumina (#E7370, New England Biolabs. DNA ends were repaired by incubation in 65 µL of 1x End Repair Reaction Buffer containing 3 µL End Prep Enzyme mix for 30 minutes at 20°C followed by 30 minutes at 65°C. A-tailing and adapter ligation were performed as follows; 15 µL of Blunt/TA Ligase master mix, 2.5 µL of freshly diluted 1.5 µM NEBNext Methylated Adapter, and 1 µL Ligation Enhancer were added, and the mixture was incubated for 15 minutes at 20°C, followed by the addition of 3 µL of USER enzyme and incubation at 37° for 15 minutes. Adapter-ligated DNA was purified using magnetic SPRI beads as follows. DNA was allowed to bind to beads (1:1 sample:bead ratio) for five minutes at room temperature, followed by collection of beads on a magnet and removal of the supernatant. Beads were washed twice with 80% (v/v) EtOH and air dried at room temperature, and DNA was eluted in 23 µL of EB.

Bisulfite conversion of adapter-ligated DNA was performed using the Qiagen Epitect kit (#59104, Qiagen) following manufacturer's recommendations for low-input/fragmented DNA. Adapter-ligated DNA was made up to 40 µL with nuclease-free water, mixed with 85 µL of Bisulfite Mix and 15 µL DNA Protect Buffer, and incubated as follows; five minutes at 95°C, 25 minutes at 60°C, five minutes at 95° C, 85 minutes at 60°C, five minutes at 95°C, and 175 minutes at 60°C. 310 µL of freshly prepared Buffer BL containing tRNA were added to bisulfite-treated DNA, and mixed by vortexing. 250 µL 100% EtOH were added, and the mixture was vortexed for 15 seconds before being loaded onto an Epitect spin column, and centrifuged for one minute at 17,000 x *g*. The column was washed with 500 µL of Buffer BW, followed by 15 minutes room temperature incubation in Buffer BD, and two further buffer BW washes. The column was centrifuged for one minute at room temperature to dry the membrane, followed by elution of DNA in 25 µL EB by centrifugation for 1 minute at 15,000 x *g*.

Final RRBS libraries were amplified by PCR as follows; 2 µL purified bisulfite-treated DNA was added to a 50 µL reaction containing 0.25 mM dNTPs, 0.5 µM NEBNext Universal Primer, 0.5 µM NEBNext Index Primer, 6.25 units *PfuTurbo* C_x HotStart DNA polymerase and 1x *PfuTurbo* C_x Reaction Buffer. PCR cycling conditions were as

follows; denaturation at 95°C for two minutes, 15 cycles of denaturation at 95°C for 30 seconds, annealing at 65°C for 30 seconds, and extension at 72°C for 45 seconds, and a final extension step of seven minutes at 72°C. Libraries were purified by two successive SPRI bead cleans as described above, but with library: bead ratios of 1:1.2 and 1:1.5. Final libraries were eluted from beads in 30 μL of EB. Library concentration was measured using a Qubit fluorometer and fragment size distribution was evaluated using an Agilent 2100 Bioanalyzer. Libraries were sequenced 100 bp paired-end on an Illumina HiSeq 2500 by the Francis Crick Institute Advanced Sequencing Facility (ASF), generating between 50 and 140 million reads per library.

2.3.3 **Data processing and analysis**

2.3.3.1 Trimming, mapping, and filtering

FastQC (Andrews, 2010) reports were inspected to confirm the quality of sequencing. Using TrimGalore! (Krueger, 2012) fastq files were trimmed for adapter content and base quality, using the command `trim_galore --paired --rrbs --retain_unpaired`. Resultant paired read fastq files were mapped to the MonDom5 (opossum) or mm10 (mouse) reference genome using Bismark (Krueger and Andrews, 2011) with the command `bismark --un --ambiguous`. Using SAMtools (Li H. et al., 2009), the bam files produced by Bismark were sorted by readname using the command `samtools sort -n` and merged into one file per sample using the command `samtools merge -n`. Methylation information at CpG positions was extracted and further processed using the R package *methylKit*. Using the function `processBismarkAln`, the methylation status of CpG sites covered at least once were extracted from bam files. In instances where paired-end reads overlapped, cytosines were counted from only the first read. Using the `filterByCoveragefunction`, data were filtered to only include CpG sites with a minimum coverage of four reads, and to exclude sites with a read coverage above the 99.9th percentile. Using the `unite` function from *methylKit*, palindromic cytosines from the Crick and Watson strands of CpG sites were combined, and data were filtered to include only those CpG sites present in at least two biological replicates of each sample condition. Finally, using the `pool` function from *methylKit*, data from individual biological replicates were merged by sample condition. The resulting filtered data objects

were saved to file using the `saveRDS` command and subsequently used for downstream analyses as detailed in sections 2.3.3.4 and 2.3.3.5 below.

2.3.3.2 CpG coverage analysis

Methylation state of sequenced cytosines were determined using the Bismark Methylation Extractor with the command `bismark_methylation_extractor --paired-end --no_overlap --gzip --multicore 2`. The genomic positions of covered CpG sites were extracted from Bismark Methylation Extractor output text files with the bash command `cut -f 3-4`. The depth of coverage was calculated as the number of occurrences of each CpG position per library using R. Line graphs showing the number of CpG sites covered at different depths were plotted using *ggplot2.*

2.3.3.3 Restriction enzyme motif analysis

Analysis of restriction enzyme motifs at insert ends was performed from untrimmed fastq files. The possible sequences resulting from digestion with MspI and BfaI followed by bisulfite conversion were identified (Table 1). The total number of reads starting with these motifs was counted using a bash script (example command `zgrep -c "^CGG"`). The percentage of reads possessing a restriction motif was calculated in R and plotted as stacked bar charts using the package *ggplot2*.

Restriction enzyme	Read 1	Read 2
MspI	CGG TGG	CAA CAG
BfaI	TAG	TAA TAG

Table 1: Possible fragment-end sequences resulting from MspI and BfaI digestion followed by bisulfite conversion.

2.3.3.4 Analysis of methylation at different genomic features

Genomic regions were defined as follows: for genes, Ensembl GTF files were transformed to `GRanges` format in R using the `import` function from the package *rtracklayer* (Lawrence et al., 2009). For the mouse, GTF version GRCm38.97 was

accessed 02-09-2019. A gene annotation file for the opossum was kindly provided by Mahesh Sangrithi, Turner laboratory. This consisted of an Ensembl GTF file (version monDom5.91, accessed 13-12-2017), to which annotation for the *RSX* locus had been added manually. Promoters were defined as the sequence 2 Kb upstream and 200 bp downstream of transcription start sites (TSS) of genes, using the function `promoters` from the R package *GenomicRanges* (Lawrence et al., 2013). Intergenic regions were defined as those bases not encompassed by a gene annotation, using the function `setDiff` from *GenomicRanges*. CGI annotation files were downloaded from UCSC (accessed 14-06-2018) and converted to `GRanges` format in R using the `readFeatureFlank` function from the package *methylKit* (Akalin et al., 2012). CGIs were defined as genic, promoter, or intergenic by their intersection with the above described genomic regions using the `subsetByOverlaps` function from the R package *IRanges* (Lawrence et al., 2013).

Filtered data objects (section 2.3.3.1) were loaded into R using the `loadRDS` command. For genes, promoters, intergenic regions, and CGIs as defined above, the methylation state of CpGs contained within each region was counted using the `regionCounts` function from *methylKit*. The `tileMethylCounts` function from the R package *methylKit* was used to construct non-overlapping 100 bp tiles, and the methylation state of CpGs contained within each tile were counted. Percentage methylation values were calculated for each counted region using the `percMethylation` function from *methylKit*. Using *ggplot2* functions, the percentage methylation data were represented as violin plots with overlaid box-and-whisker plots (to demonstrate the overall distribution of the data), or as bar charts (to display data for many individual regions). To generate smoothed line plots representing methylation at specific loci, the methylation matrices for regions of interest were extracted from the filtered data objects using `subsetByOverlaps`, and the proportion of methylated CpGs out of total CpGs was calculated per individual CpG site. These data were smoothed to loess curves using the R package *Gviz* (Hahne and Ivanek, 2016) with the `smooth` parameter to the `DataTrack` function. Graphical representations of gene annotations, CGIs, and genomic location were also generated using *Gviz*, with the `GeneRegionTrack`, `AnnotationTrack`, and `GenomeAxisTrack` functions respectively.

2.3.3.5 Differential methylation analysis

CpG sites at which methylation differed significantly between males and females were calculated using the `calculateDiffMeth` function from the R package *methylKit* (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test). The `diffMethPerChr` function was used to calculate, for each chromosome, the percentage of CpG sites with significantly different methylation. These values were plotted as stacked bars using *ggplot2*. To overlap differentially-methylated sites with different genomic features, differentially-methylated sites were extracted and intersected with gene and CGI annotations using the `getMethylDiff` and `annotateWithFeature` functions from *methylKit*.

2.4 RT-PCR analysis of Y-linked gene expression

2.4.1 PCR primer design

The opossum genome assembly does not feature an assembled Y chromosome. Therefore, in order to design RT-PCR primers against Y-chromosome genes, I made use of published opossum Y-linked coding sequences (CDS; Cortez et al., 2014, Bellott et al., 2014). RT-PCR primers ideally span a splice junction, to avoid amplification of contaminating gDNA. In order to identify splice junctions from Y-linked CDS, I first used BLAST (Altschul et al., 1990) to identify bacterial artificial chromosome (BAC) sequences containing the genomic locus from which each CDS derived. The Spidey software tool (Wheelan et al., 2001) was then used to identify exon structure from each Y-linked CDS and the corresponding BAC. Primer pairs were designed using Primer-BLAST (Ye et al., 2012), requiring the flanking of a splice junction, an amplicon size of 200-500 bp, and predicted amplicon specificity when checked against *M. domestica* RefSeq. The gene *RPS4Y* was reported as Y-linked (Cortez et al., 2014) but has previously been shown to be absent from the Y chromosome due to autosomal retrotransposition (Hughes et al., 2015). *RPS4Y* was therefore excluded from this experiment. Primers for the autosomal control gene *γ-tubulin* were designed using the same Primer-BLAST requirements. Primer sequences for the 17 Y-linked genes and the autosomal control are listed in Appendix 7.2.

2.4.2 **RNA extraction**

Two male opossums were culled using the method detailed in section 2.1. Brain, mandible, liver, spleen, testis and kidney tissues were removed, sliced into small pieces (~10 mm), and snap frozen in liquid nitrogen in Precellys 2 mL reinforced tubes (#9000538, Cayman Chemical). Tissue pieces were then stored at -80°C until use. Tissue samples were retrieved from storage on dry ice, and 1 mL TRIzol Reagent (#15596026, Invitrogen) and 2 steel lysis beads (#10400, Cayman Chemical) were added to each sample. Tissues were lysed using a Precellys 24 Homogenizer (Bertin Instruments) for two 15 second cycles at 6,500 rotations per minute (rpm), except for mandible samples, which were homogenized for two 30 second cycles at 6,500 rpm. Samples were incubated at room temperature for approximately two hours; in the case of mandible samples, the tube was held on a rocking platform to further encourage cell lysis. Samples were centrifuged at 12, 000 x *g* for 10 minutes at 4°C, and the supernatant was transferred to fresh tubes. 0.2 mL of chloroform/isoamyl alcohol was added, and tubes were shaken vigorously for 15 seconds, followed by 3 minutes room temperature incubation and centrifugation at 12,000 x *g* for 15 minutes at 4°C. The aqueous phase was transferred to fresh tubes, and 0.5 ml isopropanol was added, followed by 10 minutes room temperature incubation and centrifugation at 12,000 x *g* for 20 minutes at 4°C. The supernatant was removed, and the RNA pellets rinsed by the addition of 1 mL of 75% EtOH followed by centrifugation 12,000 x for 10 minutes at 4°C. The supernatant was removed, and pellets were left to air dry at room temperature for approximately five minutes. RNA pellets were resuspended in 50-100 µL nuclease-free H_2O and incubated at 55°C for approximately 5 minutes.

2.4.3 **cDNA synthesis**

cDNA was synthesised using the Maxima First Strand cDNA Synthesis Kit (#K1642, ThermoFisher), following manufacturer's instructions. Briefly, 1.5 µg of RNA was incubated for 2 minutes at 37°C in 10 µL of 1X dsDNase Buffer containing 1 µL of dsDNase. The mixture was adjusted by the addition of 4 µL 5X Reaction Mix, 2 µL Maxima Enzyme Mix, and 4 µL nuclease-free H_2O, and incubated for 10 minutes at 25°C. The reaction was terminated by heating at 85°C for 5 minutes, and prepared cDNA was

stored at -80°C until further use. A no-template control (NTC) was prepared in parallel. The NTC contained all reverse-transcription reagents but excluded the RNA template.

2.4.4 **PCR and gel electrophoresis**

Y-linked gene sequences were amplified from opossum male cDNA using Q5 High-Fidelity Polymerase (#M0491S, NEB) following manufacturer's instructions. Briefly, 1 µL of a 1:10 dilution of cDNA was amplified in a 20 µL reaction volume containing 1X Q5 Reaction Buffer, 0.2 mM dNTPs, 0.5 µM each of forward and reverse primers, and 0.4 units of Q5 High-Fidelity Polymerase. PCR reactions were incubated as follows; denaturation for 30 seconds at 98°C, followed by 35 cycles of denaturation for 10 seconds at 98°C, annealing for 30 seconds at an annealing temperature specific to each primerset (Appendix 7.2), and extension for 30 seconds at 72°C, and a final extension at 72°C for 2 minutes. 10 µL of each PCR product was electrophoresed through a 2% agarose-TAE gel containing 0.6 µg/mL Ethidium Bromide (EtBr) for 25 minutes at 100V, followed by visualisation of gDNA under UV light. The migration of PCR amplicons was compared to that of the GeneRuler DNA Ladder Mix (SM0331, ThermoFisher).

2.4.5 **Verification of PCR amplicons by Sanger sequencing**

PCR amplicons were purified using the NucleoSpin Gel and PCR Clean-up (#740609.250, Machery-Nagel) following manufacturer's instructions. The volume of the PCR product was adjusted to 50 µL with nuclease-free H_2O, and 100 µL of Buffer NT1 was added. PCR product was loaded onto a NucleoSpin Column followed by 30 seconds centrifugation at 11, 000 x *g*. The column was washed twice by the addition of 700 µL of Buffer NT3 followed by 30 seconds centrifugation at 11, 000 x *g.* The column was dried by centrifugation for a further 1 minute at 11, 000 x *g.* 20 µL Buffer NE was added, followed by a 1-minute incubation at room temperature. DNA was eluted by centrifugation for 1 minute at 11, 000 x *g.* Purified PCR products were submitted to Beckmann Coulter Genomics for Sanger sequencing, using the forward PCR primer as the sequencing oligo. Sequences were aligned to the *M. domestica* nucleotide collection using megablast, and in all cases aligned only to *M. domestica* Y-linked BAC sequences, and to the CDS (Cortez et al., 2014, Bellott et al., 2014) originally used for primer design.

2.5 Definition of genomic variants in opossum LL2 parental stock

In order to undertake an allele-specific analysis in opossum, it was first necessary to define the genomic variants carried by the parents of the opossum cross to be studied. To achieve this, Jasmin Zohren (Turner laboratory) re-sequenced the genomes of both parent animals and identified genomic variants in the resulting data as described briefly in the following sections.

2.5.1 Whole-genome sequencing

Ear snips from LL2 stock parent opossums ♀O9899 and ♂PO994 were prepared at the University of Texas opossum colony as part of a collaboration with John VandeBerg, and shipped to the Turner laboratory on dry ice, where they were stored at -80°C until use. gDNA was prepared using the PureLink Genomic DNA Mini kit (#K1820-02, Invitrogen). 10 mg of ear tissue was thawed on ice and macerated in a Petri dish in 100 µL PureLink digestion buffer, then transferred to a 1.5 ml microcentrifuge tube with 80 µL further PureLink digestion buffer. gDNA was then purified as described in section 2.3.1. Purified gDNA was eluted in 50 µL PureLink Genomic Elution Buffer. gDNA concentration was quantified using a Nanodrop fluorometer, and 3 μg of gDNA were submitted to the Francis Crick Institute Advanced Sequencing Facility (ASF) for library construction using the KAPA HyperPlus Kit. 150 bp paired-end sequencing was performed by the ASF on a HiSeq4000, and produced at total of 210 million reads for the male sample and 256 million reads for the female sample, with mean Phred scores of greater than 30.

2.5.2 Variant calling

Using TrimGalore! (Krueger, 2012) reads were hard trimmed at each end, as well as for adapter content and base quality, using the command `trim_galore --cores 4 --paired --fastqc --gzip --retain_unpaired --clip_R1 10 --clip_R2 10 --three_prime_clip_R1 5 --three_prime_clip_R2 5`. Libraries were mapped to the MonDom5 reference genome using BWA-MEM (Li, 2013) using the command `bwa mem -t 32 -M -R`. Paired and unpaired mapped reads were then merged to one file, sorted, and indexed using SAMtools (Li H. et al., 2009). Mapped reads were then used to call

variants using the GATK best practices pipeline (Van der Auwera et al., 2013)(Figure 8). The GATK pipeline includes a base recalibration step that typically relies on known variants as input, which were not available for opossum. The recommended solution was to iteratively call and refine variants from the same dataset. Therefore, initially variants were called using three independent pipelines; BCFtools (Narasimhan et al., 2016), Varscan (Koboldt et al., 2009), and GATK (Van der Auwera et al., 2013), and those variants identified by all three pipelines were considered to be high-confidence variants (Figure 8). The high-confidence variants were then used to perform GATK base recalibration. Subsequently, variants were called for each individual, and these variant calls were combined, and genotypes annotated to produce a variant call file (Figure 8). The complete list of 25 million variants was used to create an n-masked version of the MonDom5 reference genome using BEDtools `maskfasta`. Variants were then filtered to include only hemi- and heterozygous SNPs within each individual using a custom R script, and the resultant 2 million SNPs were formatted to a text file compatible with downstream use in the SNPsplit pipeline (Krueger and Andrews, 2016).

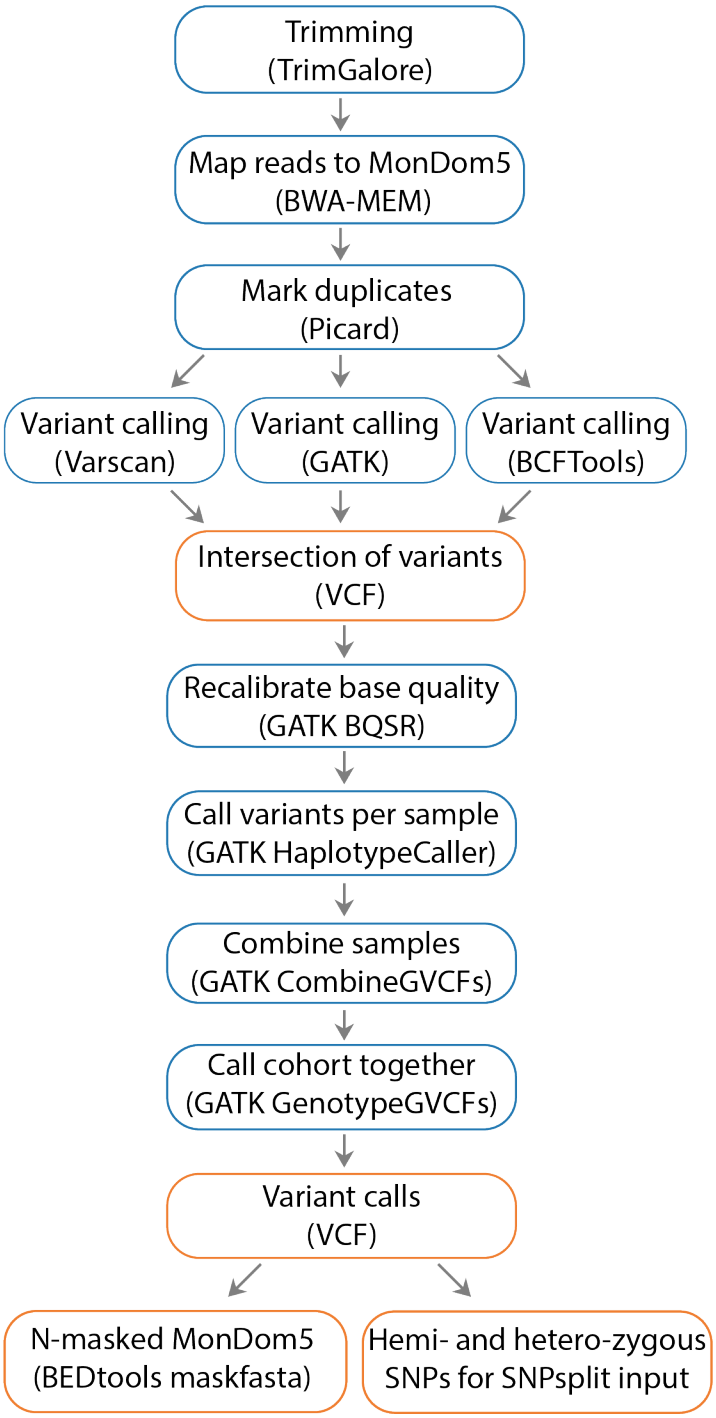


Figure 8: Outline of opossum variant calling pipeline.
Workflow diagram describing variant calling from whole genome sequencing of opossum parental genomic DNA. Blue boxes indicate processing steps, orange boxes indicate files produced. Variant calling was performed by Jasmin Zohren, Turner laboratory.

2.5.3 **Analysis of the genomic distribution of SNPs**

The genomic distribution of mouse and opossum SNPs was analysed using the `countOverlaps` function from the R package *IRanges* to count the number of SNPs occurring at different genomic features as defined in section 2.3.3.4.

2.6 Allele-specific RNA-seq

2.6.1 **RNA extraction**

RNA was prepared from mouse and opossum tissues as follows. For mouse, F1 progeny of a C57BL/6J *Xist*^{tm1Jae} x *Mus spretus* hybrid cross were culled as described in section 2.1.2.2, and brain and liver tissues were removed, cut into small pieces, and snap frozen on liquid nitrogen. Tissue pieces were stored at -80°C until further processing. Opossum brain, liver, and spleen tissues pieces were prepared from the progeny of LL2 stock opossums ♀O9899 and ♂PO994 at the University of Texas opossum colony and shipped to the Turner laboratory on dry ice, where they were stored at -80°C until use. For both mouse and opossum, tissue pieces weighing between 10-20 mg were pulverised with a pestle and mortar pre-cooled on dry ice, and RNA was prepared using the Ambion RNAqueous Micro Kit (#AM1931, ThermoFisher). Pulverised tissue was thawed on ice and 100 µL RNAqueous Lysis Buffer was added. This mixture was aspirated into a 1 mL syringe and was further homogenised by being passed through a 27 Gauge ½ ″ needle. 50 µL of 100% EtOH was added and the mixture was loaded onto an RNAqueous column and centrifuged at 17,000 x *g* for approximately 10 seconds. The column was washed by the addition of 180 µL of RNAqueous Washer Buffer 1 and centrifugation at 17,000 x *g* for approximately 10 seconds. The column was washed twice more using 180 µL room temperature RNAqueous Washer Buffer 2/3 and 40 seconds 17,000 x *g* centrifugation. The column was dried by one further minute of centrifugation at 17,000 x *g*, and was then incubated for one minute in 10 µL of RNAqueous Elution Solution pre-warmed to 75°C. RNA was eluted by 30 seconds centrifugation at 17,000 x *g,* and the elution was repeated for a total volume of 20 µL. The buffer composition was adjusted by the addition of 2 µL of 10X DNAse I Buffer, and 2 units of DNAse I were added, followed by a 20-minute incubation at 37°C. 2.3 µL of thoroughly vortexed room temperature DNAse Inactivation Reagent were added, and the reaction was incubated at room temperature for 2 minutes.

The DNAse Inactivation Reagent was pelleted by centrifugation at 17, 000 x *g* for 1.5 minutes, and the supernatant containing the purified RNA was transferred to a fresh tube. The quality of the purified RNA was confirmed using the Agilent 2100 Bioanalyzer.

2.6.2 Library preparation and sequencing

Purified RNA was submitted to the Francis Crick Institute's ASF standard RNA-seq pipeline. cDNA was prepared using the SMART-Seq v4 Ultra Low Input RNA Kit (#634894, Takara), followed by library preparation using the Nextera XT DNA Library Preparation Kit (FC-131-1096, Illumina. Library fragment size distribution was evaluated using an Agilent 2100 Bioanalyzer, and libraries were sequenced 100 bp paired-end on an Illumina HiSeq 4000. Between 54 - 156 million reads were generated per library.

2.6.3 Data processing and analysis

2.6.3.1 Trimming, mapping, and sorting of reads by allele

The quality of the raw sequencing data was confirmed by inspection of FastQC reports (Andrews, 2010). Reads were hard trimmed at each end, as well as for adapter content and base quality, using TrimGalore! with the command, `trim_galore --paired --clip_R1 10 --clip_R2 10`. A version of the mm10 reference genome in which parental genomic variants were masked with 'n' was generated using the `SNPsplit_genome_preparation` command (Krueger and Andrews, 2016), and a vcf file containing genomic variants between the C57BL/6J and *M. spretus* genomes from the Mouse Genomes Project dataset (Keane et al., 2011). The n-masked version of the monDom5 reference was generated by Jasmin Zohren, Turner lab (section 2.5). Trimmed fastq files were aligned to the n-masked MonDom5 (opossum) or mm10 (mouse) reference genomes using HISAT2 (Kim et al., 2019) with the following command, `hisat2 --no-softclip --no-mixed --no-discordant`. Using SAMtools, the resultant sam files were converted to bam format and merged by sample.

RNA-seq libraries generated from mouse tissue had mapping rates unexpectedly lower than the mapping rates for opossum libraries. To exclude that the lower mapping rates resulted from a contaminant in the mouse libraries, the unmapped reads were extracted

from the bam files, converted to fastq format, and run through FastQC (Andrews, 2010) to find any overrepresented sequences. The most overrepresented sequences were stretches of polyT. I concluded that the mouse RNA-seq libraries suffered from amplification of the polyT primer used in cDNA preparation. While this resulted in the sequencing of non-useful reads, it did not compromise the quality of those reads that did map successfully (at least ~ 40 million per library).

Using the SNPsplit tool with the command `SNPsplit --paired --conflicting`, mapped reads were flagged as deriving from either C57BL/6J or SPRET_EiJ and written to separate files according to their allelic origin. Allele-sorting statistics such as the total number of allele-specific reads, and the ratio of paternal:maternal reads were extracted from the SNPsplit output report.

2.6.3.2 Detection of XCI-escape genes

The number of reads overlapping annotated genes was counted using the `featureCounts` command from the R package *Rsubread*. Gene annotations were derived from Ensembl gtf files (GRCm38.97 and monDom5.97), and multi-mapping reads were excluded. Read counting was performed from bam files representing all mapped reads, irrespective of allelic origin, and separately for bam files representing allele-specific reads.

To determine a list of the genes that were expressed in each tissue, the `featureCounts` output for 'all mapped reads' files were merged by biological replicate, and FPKM values were calculated using the `fpkm` function of the R package *DESeq2* using the robust median ratio method. Genes were filtered to exclude gene models annotated as pseudogenes, and an expression threshold of FPKM >1 was imposed to generate a final list of expressed genes per tissue.

At genes with a low number of allele-specific reads, the allelic-expression ratio may suffer from noise (Borensztein et al., 2017). Therefore, the list of expressed genes determined above was further filtered to exclude genes with low numbers of allele-specific reads. Using the `featureCounts` output for 'allele-specific' files, the number of allele-specific reads were plotted against the number of genes retained at different

minimum read thresholds. A final gene list was generated using a threshold of ten allele-specific reads.

Subsequently, allelic expression ratios were calculated as $reads_{paternal}/(reads_{paternal} + reads_{maternal})$. As a quality control, allelic expression ratios were plotted for autosomal genes. X-linked genes with an allelic expression ratio > 0.1 were categorised as escaping XCI, following the thresholds used to identify escape genes in previous publications (Wang et al., 2014b).

2.7 Allele-specific BS-seq

2.7.1 gDNA extraction

Tissues used for preparation of gDNA were the same as those described above for RNA extraction (section 2.7.1). Following pulverisation and thawing as described in section 2.7.1, gDNA was extracted using the PureLink Genomic DNA Mini kit as described in section 2.3.1. Purified gDNA was eluted in 25 µL of PureLink Genomic Elution Buffer. The concentration of the purified gDNA was quantified using a Nanodrop fluorometer, and the high molecular weight nature of the gDNA was confirmed by electrophoresis through a 1% (w/v) agarose-TAE gel containing 0.5% (v/v) SYBR Safe (#SS33102, ThermoFisher). 3 µL of each preparation were loaded onto the gel, and current was applied at 80V for 60 minutes. 3 µL GeneRuler DNA Ladder Mix (SM0331, ThermoFisher) were run in parallel, and the gel was visualised under UV light.

2.7.2 Library preparation and sequencing

BS-seq libraries were prepared following the method of Clark *et al* (Clark et al., 2017) with the modifications suggested therein for bulk samples. 6 ng of purified sample gDNA and 6 fg of Lambda DNA spike-in (#D152A, Promega) were combined in a total volume of 10 µL EB (#19086, Qiagen). A negative control consisting of 6 fg of Lambda DNA in 10 µL EB was processed in parallel through all subsequent steps. Bisulfite conversion was carried out using reagents from the EZ Methylation Kit (#D5020, Zymo) as follows. One vial of CT Conversion Reagent was dissolved in 790 µL of M-Solubilization Buffer and 300 µL of M-Dilution Buffer by heating to 50°C with repeated vortexing, followed

by the addition of 60 µL of M-Reaction Buffer. 65 µL of prepared CT Conversion Reagent were added to each sample, and the reactions were incubated for 8 minutes at 98°C followed by 3 hours at 65°C.

Bisulfite-converted DNA was purified using the PureLink PCR Purification Column Kit (#K310050, Invitrogen) as follows: 300 µL of PureLink Binding Buffer were added to a PureLink Column and the DNA sample was transferred to the column. Using the same pipette tip, the sample tube was rinsed with 75 µL of the sample-buffer mixture, then returned to the column. Sample tubes were spun down briefly and remaining volume was transferred to the column. The column contents were briefly mixed, and centrifuged at 3,300 x *g* for 30 seconds. The column was washed using 500 µL of PureLink Wash Buffer and centrifuged at 3,300 x *g* for 30 seconds. 100 µL of M-Desulphonation Buffer (EZ Methylation Kit, Zymo) were added to the column and incubated for 15 minutes at room temperature. M-Desulphonation Buffer was removed by centrifucation at 3,300 x *g* for 30 seconds, and the column was washed twice with 300 µL PureLink Wash Buffer. The column was dried by centrifugation at 16,100 x *g* for 1 minute, and the column was incubated for 5 minutes at room temperature in 40 µL of pre-amplification mixture comprising 1 x NEB Buffer 2, 0.4 mM dNTPs, and 0.4 µM Preamp Primer (for sequence, see Appendix 7.4). DNA was eluted by centrifugation at 16,100 x *g* for 1 minute.

Samples were incubated at 65°C for 3 minutes, then immediately cooled on an ice-cold aluminium rack. Samples were briefly centrifuged, and 50 units of Klenow exo- (#M0212M, NEB) were added. The reaction was mixed by gentle vortexing and incubated at 4°C for 5 minutes in a thermocycler. The temperature was then ramped from 4 – 37°C at a ramp rate 4°C per minute, followed by a 90-minute incubation at 37°C. 40 units of Exonuclease I (#M0293L, NEB) were added, and the volume brought up to 90 µL with nuclease-free H_2O. The reactions were incubated in a thermocycler at 37°C for 1 hour with the heated lid held at 50°C. The sample volume was brought up to 100 µL with nuclease-free H_2O, combined in a fresh tube with 80 µL of AMPure XP Beads (#A63881, Beckman Coulter), and mixed by pipetting 10 times. Following a ten-minute room temperature incubation, samples were placed on a magnetic rack until the solution cleared and the beads had formed a pellet. The supernatant was removed, and the beads

were washed by pipetting twice in freshly-made 80% (v/v) EtOH, off the magnet. This wash was repeated, and subsequently the beads were air-dried at room temperature until the bead pellet lost its glossy appearance. Beads were resuspended in 49 µL of adaptor 2 tagging mixture comprising 1 x NEB Buffer 2, 0.4 mM dNTPs, and 0.4 µM Adapter 2 Oligo for NEB indices (for sequence, see Appendix 7.4).

Following a ten-minute incubation at room temperature, samples were heated to 95°C for 45 seconds, and immediately cooled on an ice-cold aluminium rack. 50 units of Klenow exo- (#M0212M, NEB) were added, and the reaction was incubated for 5 minutes at 4°C, followed by ramping from 4 – 37°C at a 4°C per minute ramp rate, and a final 90-minute incubation at 37°C. The sample volume was brought up to 100 µL with nuclease-free H_2O, 80 µL of polyethylene glycol (PEG) buffer were added, and the solution was mixed by pipetting 10 times and incubated for 10 minutes at room temperature. The mixture was placed on a magnetic rack until the beads formed a pellet, and the supernatant was removed. Beads were washed twice in 80% EtOH and air-dried as described above for previous purification steps. Beads were resuspended in 50 µL of library amplification mixture comprising 1 x KAPA HiFi Fidelity Buffer (#KK2502, KAPA Biosystems), 0.2 mM dNTPs, 0.2 µM NEBNext Universal Adaptor (#E7535S, NEB), 0.2 µM NEBNext Index Adaptor (#E7535S, NEB), and 0.02 units/µL KAPA HiFi HotStart polymerase (#KK2502, KAPA Biosystems).

Beads were incubated for 10 minutes at room temperature, followed by library amplification by PCR with the following cycling conditions: denaturing at 95°C for 2 minutes; 10 cycles of denaturing at 95°C for 80 seconds, annealing at 65°C for 30 seconds, and extension at 72°C for 30 seconds; and a final extension at 72°C for 3 minutes. Amplified libraries were brought up to 100 µL with nuclease-free H_2O, and 80 µL of PEG buffer were added, and mixed thoroughly. Samples were incubated at room temperature for 10 minutes, and subsequently placed on a magnetic rack until beads formed a pellet. The supernatant was removed, and the beads were rinsed twice with freshly-made 80% (v/v) EtOH while on the magnet. The beads were air-dried at room temperature until their glossy appearance was lost, and then removed from the magnet and resuspended in 15 µL EB (#19086, Qiagen). Following a ten-minute incubation at

room temperature, samples were returned to the magnet until the solution was clear, and the purified libraries were transferred to fresh tubes.

Library concentrations were quantified using a Qubit Fluorometer and the fragment size distribution was evaluated using an Agilent 2100 Bioanalyzer. Libraries were sequenced at 150 bp paired-end on an Illumina HiSeq 4000 by the Francis Crick Institute ASF, yielding between 198 and 363 million reads.

2.7.3 **Data processing and analysis**

2.7.3.1 Trimming and mapping

The quality of the raw sequencing data was confirmed by examination of FastQC reports (Andrews, 2010). Reads were trimmed for adapter content and base quality, and to remove the six randomly-primed bases from each end of the read, with the TrimGalore! command `trim_galore --clip_R1 6 --three_prime_clip_r1 6`. Trimmed reads were then mapped to the n-masked mouse or opossum reference genome using Bismark on single-end mode with the command `bismark --non_directional --un –ambiguous`. Library statistics such as trimming, mapping, and cytosine conversion rates were extracted from TrimGalore! and Bismark output reports.

2.7.3.2 Extraction of allele-specific methylation calls

Mapped reads were written out in different files according to their allelic origin using SNPsplit with the command `SNPsplit --bisulfite –conflicting`. Allele-specific bam files were then subject to deduplication and methylation extraction using the Bismark commands `deduplicate_bismark and bismark_methylation_extractor –bedGraph`. The total number of allele-specific reads was extracted from the SNPsplit output report.

Allele-specific CpG methylation calls were read from Bismark ‘cov’ files into R using the `methRead` function from the package *methylKit*. Data was then filtered by depth of coverage using *methylKit* commands. Autosomal and X-linked CpG sites were processed

separately, to avoid excluding X-linked sites in female samples due to low coverage of the X chromosome in male files. Using the `filterByCoverage` command, data were filtered to exclude sites in the highest 99.9 coverage percentile, and to only include sites covered at least twice in opossum, and three times in the mouse. Subsequently the `unite` function from *methylKit* was used to merge data by biological replicate, including only those CpG sites that were covered in at least two replicates for the opossum, and in all three replicates for the mouse. Data for all autosomes was extracted and saved to file. To extract methylation data for the X chromosome, data import and filtering was repeated using the same coverage parameters, but excluding the paternal genome files for male samples, as males do not carry a paternal X chromosome. Data from the X chromosome was extracted and saved to file. For subsequent analyses of methylation, the allele-specific autosomal and X chromosome datasets generated here were reloaded into R and processed and graphed as described below for specific analyses.

2.7.3.3 Global analysis of allele-specific DNA methylation

The methylation level of 100 bp non-overlapping tiles was calculated using the `tileMethylCounts` and `percMethylation` functions from the R package *methylKit*. Using the R package *ggplot2*, the overall distribution of allele-specific methylation at autosomes and the X chromosomes was represented as violin plots for each sample condition. The mean methylation value was calculated as the number of methylated CpGs divided by the total number of CpGs. Mean methylation values were superimposed on violin plots as a black diamond, and were also presented in tabular format. Smoothed line plots depicting allele-specific methylation for the X chromosome and representative autosomes were generated using *Gviz*, as described in section 2.3.3.4.

2.7.3.4 Differential methylation analysis

Data were filtered to only include female samples, and differential methylation testing was performed using the `calculateDiffMeth` command from *methylKit* to compare the paternal and maternal alleles at all covered CpG sites. The proportion of differentially methylated sites (methylation difference ≥ 25% and q-value < 0.01, Fisher's exact test) was extracted using the `diffMethPerChr` command from *methylKit*.

2.7.3.5 Analysis of allele-specific methylation at X-linked genomic features

Data were filtered to include only X-linked CpG positions. Methylation level was calculated at different genomic features using the *methylKit* functions `regionCounts` and `percMethylation`, and represented as violin plots for each sample condition. Genomic features were defined as described in section 2.3.3.4, with the addition of repeat elements as annotated by RepeatMasker (Smit et al., 2013; repeat annotation accessed 2019-10-03 and 2019-11-01 for mm10 and monDom5 genomes, respectively).

2.7.3.6 Analysis of methylation at XCI-escape genes

Methylation level at X-linked genes and promoters (calculated as in section 2.7.3.5) were represented as violin plots for male samples and for the Xa (maternal X chromosome) of female samples. For data from the Xi (paternal X chromosome) of female samples, data were categorised for plotting using the list of XCI-escape genes generated in section 2.6.3.2. In the case of the opossum, a previously published list of escape genes (Wang et al., 2014b) was incorporated into the list of escape genes. Locus-level views of representative XCI-escape and XCI-subject genes were generated using the R package *Gviz* as described in section 2.3.3.4

2.8 Low-input reduced-representation bisulfite sequencing

Low-input RRBS libraries were prepared following the method described in Guo et al. (2015b), with modifications suggested by Sebastien Smallwood (personal communication). To prevent contamination from exogenous DNA, all plasticware was UV-irradiated for one house before use, surfaces were treated with DNA*Zap* solution (#AM9890, Invitrogen) prior to use, and all steps subsequent to sample collection were performed in a dedicated pre-PCR UV hood with pipettes, plasticware, and reagents allocated exclusively for low-input library preparation.

2.8.1 **Sample preparation**

2.8.1.1 gDNA

gDNA used as a positive control in the preparation of low-input RRBS libraries was extracted from frozen tissue using the PureLink Genomic DNA Mini kit as described in section 2.3.1, and stored at 4°C. Dilutions comprising 0, 0.1, 0.5, and 100 ng/µL gDNA were used for low-input RRBS library preparation (section 2.8.2).

2.8.1.2 Mouse embryos

Following timed collection of E3.5 blastocysts as described in section 2.1.2.4, embryos were placed in Acid Tyrode's solution (#T1788, Sigma-Aldrich) and monitored under a dissecting microscope until disruption of the zona pellucida was evident. Embryos were then washed through three drops of PBS (Gibco #14190-094) and transferred to a 200 µL tube in a minimal volume of PBS. An equivalent volume of PBS from the same pipette was collected as a picking-buffer only control. Embryos and controls were snap frozen on dry ice and stored at -80°C until preparation of low-input RRBS libraries (section 2.8.2). Negative controls for each embryo were processed in parallel through the entire library preparation procedure, to control for contamination from exogenous DNA.

2.8.2 **Library preparation and sequencing**

Tubes containing embryos or picking buffer controls (section 2.8.1.2) were thawed on ice, and 1 µL of each gDNA dilution (section 2.8.1.1) was transferred to a fresh tube on ice. To these samples was added 9 units each of MspI and BfaI (#ER0541 and #ER1762, ThermoFisher) in a total volume of 18 µL 1x Tango buffer (# BY5, ThermoFisher). Samples were incubated for 3 hours at 37°C, followed by 20 minutes at 80°C. 5 units of Klenow exo- (#EP0422, ThermoFisher), and 1 µL end-repair/A-tail mixture comprising 1 mM dATP, 0.1 mM dGTP, 0.1 mM dCTP and 0.1 mM dTTP were added to each sample and incubated at 37°C for 40 minutes, followed by 75°C for 15 minutes. 30 Weiss units of T4 DNA ligase (#EL0013, ThermoFisher), 1 µL of 0.15 µM NEBNext Methylated Adaptors (#E7535S, NEB), and 0.25 µL of 50 mM ATP were added to the samples. The buffer composition was adjusted by the addition of 0.5 µL 10x Tango buffer and 2 µL nuclease-free H_2O, and adaptors were ligated by overnight incubation at 4°C. 1 µL of

USER enzyme (NEBNext Ultra II DNA Library Prep Kit for Illumina; #E7645S, NEB) was added, and the mixture was incubated at 4°C for 1 hour.

Bisulfite conversion was carried out using the Epitect Plus DNA Bisulfite Kit (#59104, Qiagen), following the manufacturer's recommendation for low-input samples. Briefly, the sample volume was made up to 40 µL with nuclease-free H_2O, and 85 µL of Bisulfite Mix and 15 µL DNA Protect Buffer were added and mixed by inversion. The reactions were incubated as follows; 95°C for 5 minutes, 60°C for 25 minutes, 95°C for 5 minutes, 60°C for 85 minutes, 95°C for 5 minutes, and 60°C for 175 minutes. 310 µL Buffer BL (containing carrier tRNA) was added, followed by 250 µL 100% (v/v) EtOH, and the sample was vortexed for 15 seconds. The sample was transferred to a MinElute Spin Column and centrifuged at 17, 000 x *g* for 1 minute. The column was washed by the addition of 500 µL Buffer BW followed by centrifugation at 17, 000 x *g* for 1 minute. 500 µL Buffer BD were added, and column was incubated at room temperature for 15 minutes, followed by centrifugation at 17, 000 x *g* for 1 minute. The column was washed twice with 500 µL Buffer BW, and once with 250 µL of 100 % (v/v) EtOH. The column membrane was dried by centrifuged at 17, 000 x *g* for 1 minute. 30 µL of EB pre-warmed to 60°C was added and the column was incubated at room temperature for 1 minute. The sample was eluted by centrifugation at 15, 000 x *g* for 1 minute.

Bisulfite-converted DNA was amplified by PCR in a 50 µL reaction comprising 0.2 mM dNTPs, 0.2 µM each NEBNext Universal Primer and NEBNext Index Primer (#E7535S, NEB), and 1 unit *PfuTurbo Cx* HotStart DNA Polymerase (#600412, Agilent) in 1x *PfuTurbo Cx* Buffer. Reactions were incubated in a thermocycler as follows. DNA was denatured at 95°C for 2 minutes, followed by cycles of denaturation at 95°C for 20 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 1 minute, and a final extension step at 72°C for 5 minutes. Libraries prepared from gDNA dilutions were amplified by 24 cycles of PCR; embryo libraries by 20 or 24 cycles. 90 µL of room-temperature SPRI beads were added to the PCR-amplified libraries, mixed by pipetting, and incubated at room temperature for 15 minutes. Samples were placed on a magnetic rack until the solution was clear, the supernatant was removed, and the beads were washed twice with freshly-made 80% (v/v) EtOH. Beads were air-dried until they lost their glossy

appearance, and then resuspended in 18 µL EB (#19086, Qiagen) and incubated at room temperature for 2 minutes. Samples were transferred to the magnetic rack until the supernatant was clear, and a 15 µL volume was transferred to fresh tubes and stored at -20°C until use. Library concentrations were evaluated using a Qubit Fluorometer, and the fragment size distribution was checked using an Agilent 2100 Bioanalyzer.

2.8.3 **Data processing and analysis**

2.8.3.1 Trimming, mapping, and extraction of methylation calls

Raw sequencing files were inspected using FastQC and trimmed for sequencing quality and adapter contamination using TrimGalore! with the command `trim_galore --rrbs --paired --retain_unpaired`. Trimmed reads were re-examined using FastQC to verify that cytosine content and adapter contamination issues were resolved, before mapping to the mouse reference genome (mm10) using Bismark with the command `bismark --un –ambiguous`. Methylation calls were extracted using the Bismark Methylation Extractor with the command `bismark_methylation_extractor --paired-end --no_overlap`.

2.8.3.2 CpG coverage and restriction enzyme motif analyses

Analysis of CpG coverage and restriction enzyme motifs at read ends were carried out as described for high-input RRBS in sections 2.3.3.2 and 2.3.3.3.

2.8.3.3 Analysis of methylation at different genomic features

Methylation calls were imported into Seqmonk (Andrews, 2007) and filtered to exclude sites covered by fewer than five reads. Using Seqmonk, genomic probe sets representing 5-kb non-overlapping windows, gene bodies, CGIs, and imprint-associated DMRs were generated for the mouse genome. The Seqmonk Running Window Probe Generator function was used to define 5-kb windows, and the Feature Probe Generator function was used to generate gene, CGI, imprinted DMR probe sets. Gene and CGI annotations were the same as those used for high-input RRBS in section 2.3.3. Imprinted DMR annotations were downloaded from the Web Atlas of Murine Imprinting and Differential Expression (WAMIDEX) (Schulz et al., 2008), and transferred from mm9 to mm10 genomic

coordinates using the USCS Genome Browser liftOver tool (Kent et al., 2002). Methylation levels were calculated at each probe set using the Seqmonk Bisulfite Methylation Over Features function. Methylation levels at features were saved to file, imported into R, and represented as violin plots or bar charts using *ggplot2*.

2.9 Low-input BS-seq

Low-input BS-seq libraries were prepared following the method detailed in Clark *et al* (2017). For samples comprising more than one cell, i.e. embryos and pooled sperm, the protocol was modified as recommended for low-bulk samples. For all sperm samples, the protocol was further modified to include harsher cell lysis and lengthened bisulfite conversion, as recommended for preparation of sperm libraries (Stephen Clark, personal communication). To prevent contamination with exogenous DNA, libraries were prepared in a dedicated pre-PCR workspace as described for low-input RRBS in section 2.8.

2.9.1 Sample preparation

2.9.1.1 gDNA

gDNA was extracted from frozen tissue using the PureLink Genomic DNA Mini kit as described in section 2.3.1 and used as a positive control for optimisation of low-input BS-seq library preparation.

2.9.1.2 Mouse embryos

E3.5 mouse embryos were collected as described above for low-input RRBS (section 2.8.1) and used in optimisations of low-input BS-seq libraries.

2.9.1.3 Opossum embryos and oocytes

Following timed collection as described in section 2.1.1.4, opossum embryos and oocytes were processed to remove the surrounding shell coat and mucoid layer, to ensure sperm and maternal somatic cells trapped therein were not included in final samples. Using a

dissecting microscope, with the sample under PBS, the shell coat was gently nicked with 1 µm tip dissecting needles (#10130-20, FST). The sample was incubated in 5 mg/mL protease in PBS (#P8811-100MG, Sigma) at 32°C for between 2 and 7 minutes, checking periodically under the dissecting microscope until the mucoid coat had clearly begun to break down. Protease digestion was halted by transferring the sample back into fresh PBS. Dark-field illumination was used to visualise any remaining mucoid coat, which was removed by gentle disaggregation with dissecting needles. The sample was then washed through a minimum of three droplets of fresh PBS and transferred to a 200 µL tube in minimal PBS, before being snap frozen on dry ice and stored at -80°C until further processing. To control for contamination from exogenous DNA, a similar volume of PBS from the same pipette was collected as a picking-buffer only control, and processed in parallel through the entire library preparation procedure.

2.9.1.4 Sperm

Male opossums and mice were culled as described in section 2.1, and epididymides were dissected from the testes and rinsed in PBS. The cauda epididymides were removed and placed in a small plastic dish containing 2 mL Bigger-Whitten-Whittingham (BWW) Buffer (opossums) or TYH Buffer (mice). Several small incisions were made in each piece of tissue, and epididymides were incubated at 37°C for 30 minutes to facilitate sperm swim out.

2.9.1.4.1 PureSperm Gradient

Purification of opossum and mouse sperm using a PureSperm gradient was performed as follows. Solutions of 90% and 45% (v/v) PureSperm 100 (#PS-100, Nidacon) were prepared by dilution in PureSperm Buffer (#PSB-100, Nidacon). 2 mL of 90% solution were added to a 15 mL Falcon tube, and 2 mL of 45% solution were layered on top to create a gradient. 1.5 mL of sperm swim-out suspension were transferred on top of the gradient, and the remainder was kept on ice for later quantification. The gradient was centrifuged for 20 minutes at 300 x *g* with no brake, and the supernatant was removed, taking care not to disrupt the sperm pellet by leaving approximately 4 mm of the 90% solution undisturbed. The pellet was resuspended in 5 mL PureSperm Wash Buffer

(#PSW-100, Nidacon), and the solution was centrifuged for 10 minutes at 500 x *g* with no brake. The supernatant was aspirated, leaving minimal liquid above the pellet, except in cases where no pellet was visible, where 0.25 mL of liquid was left. The purified sperm was resuspended in 300 µL PBS and kept on ice for subsequent quantification.

2.9.1.4.2 Differential lysis

Purification of opossum and mouse sperm from contaminating somatic cells by differential lysis was performed as follows. 250 µL of sperm swim-out suspension was diluted with 250 µL 1x SSC containing either 0.1%, 0.5%, or 1% (v/v) SDS. Control samples were diluted in 1x SSC with no SDS. Suspensions were mixing by pipetting to encourage cell lysis. The suspension was kept on ice until quantification.

2.9.1.4.3 Sperm purity quantification

The percentage of sperm and somatic cells was assessed by manually counting cells using a glass haemocytometer with improved Neubauer rulings. A square coverslip was affixed to the haemocytometer by two small water droplets. The ice-chilled cell suspensions were mixed by gentle pipetting and 10 µL transferred under each side of the coverslip by capillary action. Cells were left to settle for 15 minutes in a humid chamber, then counted using a compound light microscope with a 20x or 40x objective. A minimum of 100 cells were counted, except in the instances where neither somatic nor sperm cells could be observed. In these cases, the entirety of both haemocytometer grids was checked before concluding that no cells had endured the purification procedure.

2.9.1.4.4 Manual picking of sperm cells

For manual picking of sperm cells, the sperm swim-out suspension was diluted approximately 1:10 in PBS, and examined under a dissecting microscope using a dark field. A field of view was located in which only one sperm cell could be seen, and a Stripper pipette with a 100 µM tip (MXL3-STR and MXL3-100, Cooper Surgical) was used to aspirate an individual sperm. Each sperm cell was then washed through a minimum of three fresh PBS drops, before being deposited in a 200 µL tube containing

5 µL RLT Plus Buffer (#1053393, Qiagen) containing 2% (v/v) ß-mercaptoethanol. Tubes were snap frozen on dry ice and stored at -80°C until use. For collection of pools of sperm, multiple sperm cells were picked in one pipette and processed together through washing, collection, and freezing as for single sperm cells. Picking-buffer only controls were collected as described above for embryo collections, and processed in parallel through the library preparation procedure.

2.9.2 **Library preparation and sequencing**

Embryo and oocyte samples, along with their respective picking-buffer controls, were thawed on ice, and 2.5 µL of RLT Plus Buffer (#1053393, Qiagen) was added, along with 7.5 µL of EB containing 6 fg Lambda DNA spike-in (#D152A, Promega). A negative control comprising 6 fg Lambda DNA in 10 µL EB was prepared in parallel and carried through all subsequent library preparation steps. Bisulfite conversion was carried out using the EZ Methylation Kit (#D5020, Zymo) as described in section 2.7.2. Sperm samples were thawed on ice, 5 µL Sperm Lysis Buffer were added, and samples were incubated at 50°C for 6 hours. Subsequently, 65 µL of CT Conversion Reagent (EZ Methylation Kit; #D5020, Zymo) were added, and bisulfite conversion was completed with the following modified incubation conditions. 98°C for 8 minutes, 64°C for 30 minutes, 98°C for 3 minutes, and 64°C for 3 hours.

Purification of bisulfite- converted DNA was carried out using the EZ Methylation Kit (#D5020, Zymo) and the PureLink PCR Purification Column Kit (#K310050, Invitrogen) as described in section 2.7.2. The remaining library preparation steps, including preamplification, exonuclease treatment, adaptor 2 tagging, and SPRI bead purification of adaptor-ligated DNA were carried out as described for high-input BS-seq in section 2.7.2. Oocytes and individual sperm cells underwent a modified preamplification step recommended for single cells as follows. Following elution of bisulfite-converted DNA into preamplification mixture, samples were heated to 65°C for 3 minutes, immediately cooled on an ice-cold aluminium rack, and incubated at 4°C for 5 minutes, ramped from 4°C to 37°C at 4°C per minute, and incubated at 37°C for 30 minutes. Samples were then heated to 95°C for 45 seconds, immediately cooled on an ice-cold aluminium rack, and topped up with 2.5 µL of preamplification mixture comprising 1x NEB Buffer 2, 0.4 mM

dNTPs, 4 μM Preamp Oligo, and 25 units Klenow exo- (#M0212M, NEB). The reaction was then incubated as described above, and the reagent top-up and incubation steps were repeated a further three times, to result in a total of 5 rounds of preamplification. For the final incubation, the 37°C step was extended to 90 minutes.

For all samples, library amplification PCRs were carried out as described for high-input BS-seq in in section 2.7.2, but with varying numbers of PCR cycle depending on the number of cells in the input sample (Table 2). Amplified libraries were purified using SPRI beads as described in section 2.7.2, and the concentration and size distribution were assessed using a Qubit Fluorometer and 2100 Bioanalyzer respectively. Libraries were sequenced on an Illumina HiSeq 4000 by the Francis Crick Institute ASF, attaining between 70 and 130 million reads per library.

	Sample	**Number of PCR cycles**
Mouse	Dilute gDNA	14
	E3.5 blastocysts	14
Opossum	Sperm	14 -18
	Oocytes	19
	Opossum E1.5 to E5.5 embryos	19
	Opossum E6.5 and E7.5 embryos	10-14

Table 2: PCR cycle conditions for low-input BS-seq

2.9.3 Data processing and analysis

2.9.3.1 Trimming, mapping, and extraction of methylation calls

Quality control of raw sequencing data was carried out using FastQC, and fastq files were concatenated to generate one file per individual library. Using TrimGalore! with the command `trim_galore --clip_R1 6 --three_prime_clip_r1 6`, fastq files were trimmed in single-end mode to remove the 6 bases derived from random-priming oligonucleotides, as well as to remove low quality bases and adapter sequences. Data were mapped to reference genome using Bismark with the command `bismark --non_directional --un –ambiguous`. For mouse gDNA and E3.5 blastocyst samples, the mm10 reference genome was used. Opossum samples were mapped to a version of the monDom5 reference genome modified to include a pseudoY chromosome comprising previously reported Y-linked CDS (Cortez et al., 2014, Bellott et al., 2014) interspersed with Ns. Reads were deduplicated using the command `deduplicate_bismark –bam` and

methylation calls were extracted using the command `bismark_methylation_extractor --bedGraph`. Library statistics including trimming, mapping, and cytosine conversion rates were extracted from TrimGalore! and Bismark output reports.

2.9.3.2 CpG coverage analysis

The number of CpG sites captured per library was calculated using the approach described in section 2.3.3.2. For mouse low-input BS-seq libraries, the absolute coverage values were plotted using *ggplot2*, whereas for opossum libraries the minimum, maximum, and average CpG coverage values were plotted for each timepoint.

2.9.3.3 Principle components analysis

For principle components analysis (PCA) in both mouse and opossum, CpG methylation calls were imported into R from Bismark 'cov' files using the `methRead` function from the package *methylKit*. Using the *methylKit* `PCASamples` function, a PCA was performed on methylation calls for all covered CpG sites. Principle components one and two were plotted using the `autoplot` function from the package *ggfortify*.

2.9.3.4 Analysis of methylation in optimisation samples

CpG methylation calls were imported from Bismark 'cov' files into R using the `methRead` function from the package *methylKit.* No coverage filtering was imposed.
Methylation levels were quantified at 100 bp non-overlapping genomic tiles using the *methylKit* functions `tileMethylCounts` and `percMethylation`. Genomic features including genes, CGIs, and imprinted DMRs were defined as described in section 2.8.3.3, and methylation levels were quantified at these features using the *methylKit* functions `regionCounts` and `percMethylation`. Methylation levels were represented as violin plots or heatmaps using *ggplot2.*

2.9.3.5 Categorisation of opossum embryo samples as male or female

The sex of embryo samples was identified on the basis of the number of reads mapping to the sex chromosomes. Using SAMtools, bam files were sorted and indexed using the

commands `samtools sort` and `samtools index`. The number of reads mapping to the X and pseudoY chromosomes were counted from bam files (example command: `samtools view -c -q 1 file.bam 'chrPseudoY'`). Data were imported into R, converted to a percentage of total mapped reads, and represented as scatterplot using *ggplot2*. Upon visual inspection, thresholds were chosen to categorise samples as male or female, and dashed lines were overlaid on the scatterplot at $y = 0.0002$ and $x = 2.2$ to indicate the chosen thresholds. Embryo samples with <0.0002 % reads mapping to the pseudoY were categorised as female, and those with >0.0002 % pseudoY-mapped reads were categorised as male.

2.9.3.6 Analysis of global methylation in opossum embryo samples

2.9.3.6.1 Data import and filtering

For analysis of global methylation in opossum early development, data from individual libraries were combined to generate one *in silico* sample for each timepoint. CpG methylation calls were read into R from Bismark 'cov' files separately for each timepoint, using the `methRead` function from the package *methylKit*. Data were then pooled using the commands `unite` (with the parameter `min.per.group = 0L)` and `pool` from *methylKit*. *In silico* pooled datasets for each timepoint were extracted as dataframes using the *methylKit* command `getData` and filtered to include only CpG sites with a coverage of >2 reads. For each CpG site, a methylation ratio (number of methylated sites/depth of coverage) was computed and added to the dataframe, which was saved to disk. Data for all timepoints were subsequently imported into R using the `methRead` function with the parameter `pipeline = list`. The data were filtered to include only CpG sites covered in all timepoints, thereby generating a consensus set of CpG sites captured at all timepoints with a minimum coverage of three reads. For comparison to methylation in adult opossum tissues, high-input BS-seq data generated in section 2.7 were imported into R using the command `methRead` and converted to a `GRanges` object. A `GRanges` object was generated representing the consensus CpG positions covered in low-input libraries, and the adult data were filtered to include only the consensus sites using the command `subsetByOverlaps`. This dataset was then used to analyse global features of methylation as detailed in the following sections.

2.9.3.6.2 Analysis of genomic distribution of covered CpG sites

The location of the consensus CpG sites was compared with genomic features including genes, promoters, CGIs, and repeat elements using the `findOverlaps` function, and the percentage of CpGs overlapping in each feature was graphed using *ggplot2*. Genomic features were defined in sections 2.3.3.4 and 2.7.3.5. The RepeatMasker annotation was manually recoded to group repeats into five broad classes (DNA, LINE, SINE, LTR, and RNA).

2.9.3.6.3 Analysis of global methylation levels

The methylation level for each individual CpG site and over different genomic features was calculated using the `percMethylation` function. Genomic features were defined as described in section 2.7.3.5. The overall methylation distribution per timepoint was represented as violin plots using *ggplot2*, and the heatmaps showing the methylation level per CpG site were generated using the function `Heatmap` from the R package *ComplexHeatmap* (Gu et al., 2016). The heatmap columns were manually ordered according to sample timepoint, while the heatmap rows was allowed to cluster based on Euclidean distances using the *hclust* method `complete`.

2.9.3.7 Differential methylation analysis

For analysis of differential methylation between each successive timepoint of opossum early development, data from individual libraries were combined to generate one *in silico* sample for each timepoint as described in section 2.9.3.6.1 above. *In silico* pooled data representing each timepoint were reimported into R using the `methRead` function with the parameter `pipeline = list`. Using the command `unite`, a filtered dataset was generated comparison containing only CpG sites represented in both timepoints for each successive timepoint. Sites of differential methylation between timepoint comparisons (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test) were calculated using the `calculateDiffMeth` and `diffMethPerChr` commands. The number of hyper- and hypomethylated sites were extracted and plotted as stacked bar charts for each timepoint

comparison using *ggplot2*. Using the `annotateWithFeature` function, the proportion of differentially methylated regions between sperm and oocytes that overlapped different genomic features were annotated and plotted as bar charts using *ggplot2*. Genomic features were defined as described in section 2.7.3.5.

2.9.3.8 Analysis of X-chromosome methylation

Data were imported and filtered following the procedure described in section 2.9.3.6.1, but with samples *in silico* pooled separately by sex at each timepoint. For analysis of bulk methylation levels, CpGs were filtered to include all sites covered by three or more reads within each sample condition but were not filtered for coverage between samples, i.e. CpG sites were not required to be covered in all timepoints. Methylation level was calculated per CpG site as the percentage of methylated CpGs/ covered CpGs, and the mean of these values was taken as the global methylation level for autosomes, chrUn, and the X chromosome and plotted using *ggplot2*.

For analysis of methylation levels at genomic features, the male and female dataset for each timepoint were further filtered using the `unite` function, i.e. for each timepoint, male and female *in silico* samples contained the same CpG positions. Methylation levels were calculated and plotted for different genomic features as described in section 2.9.3.6.3.

Locus-level plots representing *RSX* and the autosomal control gene *RUNX1* were generated using *Gviz* as described in section 2.3.3.4. Data were subject to the same filtering as described for bulk methylation levels, i.e. for a particular sample, all those CpG sites covered by at least three reads were included. For the *RSX* locus, custom CGI annotations were generated by Fanny Decarpentrie (Turner laboratory) using the DBCAT tool (Kuo et al., 2011).

2.10 Analysis of gene expression in opossum embryos

Single-cell RNA-seq data representing opossum oocytes and E1.5-7.5 embryos were kindly shared with me by Shantha Mahadevaiah and Mahesh Sangrithi (Turner

laboratory), who generated this dataset as part of an unpublished project. Briefly, opossum embryos were collected following a similar procedure to that described in sections 2.1.1.4 and 2.9.1.3, and embryos were manually disaggregated for single cell collection. RNA-seq libraries were prepared using the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (#634894, Clontech) and the Ovation Ultralow Library system V2 1-96 (# 0347, NuGen), and sequenced 100 bp paired-end on an Illumina HiSeq 4000 by the Francis Crick Institute ASF. Sequences were mapped to the monDom5 reference genome using HISAT2, and StringTie (Pertea et al., 2015) was used to calculate gene-level transcript abundances. The resulting expression matrix was made available for my analysis of the expression of genes involved in DNA methylation pathways. Opossum DNA methylation genes were identified based on their orthology to mouse or human DNA methylation factors using the Ensembl database, and $\log_2$(FPKM+1) gene expression values were plotted using *ggplot2*.

2.11 Immunofluorescent staining

2.11.1 Paraffin sections

Ovaries and testes were dissected from recently culled adult mouse or opossum, rinsed in PBS, and fixed overnight in 4% PFA in PBS at 4°C. The following day, tissues were washed for 30 minutes in PBS at room temperature and transferred to 70% EtOH. Tissue samples were paraffin embedded and sectioned at 6 μm by the Francis Crick Institute Experimental Histopathology facility. Paraffin was removed by baking slides at 60°C for one hour followed by two one-minute incubations in 100% xylene. Sections were rehydrated through an EtOH gradient (two times one minute in 100% EtOH, one minute in 70% EtOH, and one minutes in MilliQ H_2O). Antigen retrieval was performed by boiling slides for 10 minutes in 0.1 M sodium citrate with replacement of any lost liquid level with MilliQ H_2O, followed by cooling of slides for 20 minutes. Slides were rinsed in PBS-T (PBS + 0.1% Triton X-100), and incubated in 3.5 M HCL for 15 minutes at room temperature in a glass coplin jar. Slides were rinsed in PBS-T and blocked for 20 minutes in blocking buffer (5% FBS in PBS-T). Slides were placed in a humid chamber and incubated with primary antibodies (mouse anti-5mC; #33D3, Epigentek and rabbit anti-DDX4; #ab13840, Abcam) at 1:100 in blocking buffer overnight at 4°C. Slides were washed three times for five minutes in PBS-T and incubated with secondary antibodies

(donkey anti-mouse AlexaFluor-568, A10037, ThermoFisher and chicken anti-rabbit AlexaFluor-488, A21441, ThermoFisher; 1:100 in PBS-T) in a humid chamber for 45 minutes at room temperature. Slides were washed three times for five minutes in PBS-T at room temperature, protected from light. Slides were mounted with VectaShield (ThermoFisher) and imaged using an Olympus IX70 inverted microscope with an Olympus UPlanApo 60×/1.40 oil-immersion objective. Captured images were processed using Fiji software to adjust brightness and contrast and prepare false-colour merge images.

2.11.2 **Whole mount oocytes and zygotes**

Opossum zygotes were collected from the uterus at 26 and 29 hpc as detailed in section 2.1.1.4. Oocytes were retrieved at approximately 24 hpc by manual disaggregation of opossum ovaries under PBS using one micron-tipped dissection needles. Oocytes were then washed a minimum of three times through clean PBS. Zygote immunostainings were performed by Sugako Ogushi, Turner laboratory. Samples were washed repeatedly in PBS and fixed for one hour on ice in 4% PFA (freshly prepared in PBS, passed through a 0.22 μM filter, and pre-cooled to 4°C). Samples were washed three times in PBS-T (PBS+0.1% Triton X-100), followed by permeabilization for 20 minutes in PBS + 0.5% Triton X-100, and in the case of oocytes, overnight storage in PBS+0.2% Triton X-100 at 4°C. Samples were incubated in 3.5 M HCL for 20 minutes at room temperature in glass-bottomed dishes (#D35-14-1.5N, Cellvis), followed by three changes of PBS-T and one 40-minute wash in PBS-T. Samples were incubated for one-four hours in blocking solution (10% FBS in PBS-T, freshly prepared and passed through a 0.22 μM filter). Samples were incubated in primary antibody (mouse anti-5mC; #33D3, Epigentek, rabbit anti-H3K9me3; #07-442, Millipore) 1:100 in blocking solution at 4 °C overnight. The following day, samples were washed quickly by three changes of PBS-T followed by one 20-minute wash in PBS-T. Samples were incubated in secondary antibody (chicken anti-mouse AlexFluor-488, #A21200, ThermoFisher, chicken anti-rabbit AlexaFluor-594, #A21442, ThermoFisher) at 1:300 in PBS-T for 1 hour at room temperature protected from light. Samples were washed quickly by three changes of PBS-T followed by one 20-minute wash in PBS-T. Oocyte samples were mounted in a 1:30 dilution of VectaShield (#H1000, ThermoFisher) in PBS in a glass-bottomed dish (#D35-14-1.5N,

Cellvis) and imaged using an Olympus IX70 inverted microscope. Zygote samples were incubated in 100 µg/ml propidium iodide (PI) for 20 minutes at room temperature, followed by two changes of PBS-T and one 30-minute PBS-T wash. Zygotes were mounted in VectaShield on Superfrost™ Microscope Slides (#AA00008032E00MNT10, ThermoFisher), under 18x18 mm coverslips (#474030-9000-000, Zeiss) held up by silicone grease (#Z273554-1EA, Sigma) and sealed with clear nail polish. Images were captured using a Zeiss Invert710 confocal microscope. Captured images were processed using Fiji software to adjust brightness and contrast and prepare false-colour maximum intensity projection images.

Quantification of anti-5mC and anti-H3K9me3 staining in the pronuclei of zygotes was carried out by Sugako Ogushi (Turner laboratory), as follows. Briefly, captured images were loaded into ImageJ, and converted to black and white. To identify the areas corresponding to the two parental pronuclei, the images were 3D segmented using the PI channel. The average pixel intensity in the segmented areas was quantified for PI, anti-5mC and anti-H3K9me3, and the 5mC and H3K9me3 values were normalised to PI. The quantification data were imported into R, and the difference in signal intensity between pronuclei (pn) was calculated for each zygote as pn1-pn2/ (pn1 + pn2)/2, and graphed using *ggplot2*. Paired measures for 5mC and H3K9me3 from the same zygote are indicated by the colour of each data point.

Chapter 3. Results 1

In vertebrate genomes, methylation is largely observed at cytosines in the CpG context, apart from a low level of non-CpG methylation in embryonic stem cells, oocytes, and neurons (Lister et al., 2009a, Tomizawa et al., 2011, Lister et al., 2013). CpGs are underrepresented in vertebrate genomes, with the exception of the regions of densely clustered CpGs termed CpG islands (CGIs). Vertebrate genomes are largely hypermethylated across intergenic and intragenic regions, but are interspersed with regions of low methylation at CGIs (Bird, 1986).

In mammals, DNA methylation is required for successful development (Li et al., 1992a) (Okano et al., 1999). The suggested functions of DNA methylation differ depending on genomic context. The requirement to control the expression of the large number of transposable elements (up to 50% of mammalian genomes) is suggested as the reason for the global hypermethylation of the vertebrate genome (Yoder et al., 1997, Walsh et al., 1998). Methylation is observed across gene bodies, and appears to be positively correlated with active transcription (Lister et al., 2009a). The functions of gene body methylation are not yet firmly established, but it is suggested that it may repress cryptic intragenic promoters (Neri et al., 2017), or play a role in active transcription (Yang et al., 2014). The position of CGIs is strongly correlated with the location of gene promoters, and while most CGIs are hypomethylated across tissues, specific examples of gene repression as a result of hypermethylation at CGI promoters have been the classical example of gene regulation by DNA methylation. For example, in the repression of certain tissue-specific and germline genes (McGhee and Ginder, 1979, Jones and Taylor, 1980, Borgel et al., 2010, Velasco et al., 2010, Hackett et al., 2012), in the regulation of allele-specific expression at imprinted loci (Stoger et al., 1993, Ferguson-Smith et al., 1993, Li et al., 1993, Bartolomei et al., 1993), and in the silencing of genes on the inactive X chromosome in females.

The phenomenon of X inactivation in mammals occurs in order to achieve dosage compensation between female (XX) and male (XY) mammals. The mammalian

XY sex chromosomes evolved from an ancestral autosome pair before the divergence of metatherian and eutherian mammals. The acquisition of the testis-determining gene *Sry* on one member of the ancestral pair formed the proto-Y chromosome (Foster and Graves, 1994). Subsequent inversions led to the suppression of recombination between the pair, leaving the proto-Y vulnerable to gene loss via genetic drift (Lahn and Page, 1999). The loss of most ancestral genes from the Y chromosome resulted in the modern large X- and small Y chromosomes, producing a dosage imbalance for X-linked gene products.

X chromosome inactivation (XCI) is mediated in eutherians by the long non-coding RNA (lncRNA) *Xist* (Brockdorff et al., 1991, Borsani et al., 1991, Brown et al., 1991a, Penny et al., 1996, Marahrens et al., 1997). *Xist* is transcribed from the future inactive X (Xi), which is chosen at random in the early embryo, and spreads along the chromosome in *cis,* initiating the recruitment of repressive epigenetic marks and establishing a silent state.

Female metatherian mammals also silence one X chromosome to achieve dosage compensation. However, unlike eutherians, XCI is imprinted, with the paternal X always being silenced. There is no *Xist* homologue in metatherians (Duret et al., 2006, Hore et al., 2007, Shevchenko et al., 2007); rather, the lncRNA *RSX* is suggested to play a similar role as *Xist* does in eutheria. *RSX* is expressed exclusively in female cells and coats the Xi in *cis*, suggesting that metatherians and eutherians evolved convergent mechanisms of XCI (Grant et al., 2012). Some XCI-associated chromatin characteristics are conserved between eutherians and metatherians, for example H3K27me3 coating of the silent X (Mahadevaiah et al., 2009).

In eutherians, a canonical example of XCI-associated repressive epigenetic marks is the hypermethylation of CGIs on the Xi. This pattern was observed in early single-locus methylation studies in both mouse (Lock et al., 1986, Singer-Sam et al., 1990, Norris et al., 1991, Grant et al., 1992, Blewitt et al., 2008) and human (Wolf et al., 1984a, Wolf et al., 1984b) (Keith et al., 1986, Tribioli et al., 1992). The repressive nature of Xi CGI methylation was confirmed by studies showing CGI methylation to be lost upon reactivation of Xi genes by 5-azacytidine treatment (Yen et al., 1984, Pfeifer et al., 1990), and necessary for the maintenance of the silent state in embryonic lineages (Sado et al.,

2000). With the advent of chromosome-wide technologies for the profiling of DNA methylation, the observation of Xi hypermethylation was confirmed to extend to X-linked CGIs across the chromosome in mouse (Gendrel et al., 2012, Auclair et al., 2014) and human (Yasukochi et al., 2010, Sharp et al., 2011, Cotton et al., 2011, Joo et al., 2014, Cotton et al., 2015).

Some early studies also examined X-chromosome methylation outside of CGIs, and found evidence for hypermethylation of the active X (Xa), both in single-locus profiles of inter- or intragenic regions in human (Wolf et al., 1984b, Lindsay et al., 1985, Boyd and Fraser, 1990, Giacalone et al., 1992) and mouse (Lock et al., 1986), and using labelling-based techniques to profile whole human chromosomes (Viegas-Pequignot et al., 1988), though different labelling techniques gave the opposite result (Miller et al., 1982, Prantera and Ferraro, 1990).

The ability to profile DNA methylation at a chromosomal level allowed the examination of methylation profiles at non-CGI regions, prompting the interesting finding that gene bodies on the human Xa are highly methylated, presumably in correlation with active gene expression (Hellman and Chess, 2007). Other findings suggest that in addition to genic regions, intergenic regions are also hypermethylated on the human (Weber et al., 2005) and mouse Xa (Gendrel et al., 2012, Keown et al., 2017, Duncan et al., 2018). An exception to this pattern occurs at genes that escape from, or are not subject to, XCI; at these loci, promoter-associated CGIs are equivalently hypomethylated on the Xi and Xa, and gene bodies equivalently hypermethylated on the Xa and Xi (Goodfellow et al., 1988, Norris et al., 1991, Blewitt et al., 2008, Cotton et al., 2011, Sharp et al., 2011, Schultz et al., 2015).

Together, these methylation data suggest a scenario wherein the Xa is hypermethylated relative to the Xi, with the notable exception being CGIs, where the Xi is hypermethylated, and genes escaping XCI, where individual loci resemble the Xa. In this scenario, the Xa resembles a typical chromosome, methylated at inter- and intra-genic regions, and hypomethylated at CGIs. In contrast, it is methylation on the Xi that deviates from the typical genomic pattern.

Past studies have suggested that the metatherian Xi does not display CGI hypermethylation. Four studies, examining a total of 29 X-linked promoters, reported that Xi promoters were unmethylated (Kaslow and Migeon, 1987, Loebel and Johnston, 1996, Wang et al., 2014b). A further two studies, using methyl-sensitive restriction digestion and immunofluorescence, reported global hypomethylation of the Xi (Loebel and Johnston, 1993, Rens et al., 2010). A recent study has extended these findings using reduced-representation bisulfite sequencing (RRBS), confirming that promoters are not hypermethylated on the metatherian Xi, and suggesting that regions flanking the promoter may be differentially methylated between the Xa and Xi (Waters et al., 2018). However, the methylation landscape at base-pair resolution remains to be established for regions outside of CGIs/promoters.

In this chapter, I apply reduced-representation bisulfite sequencing to brain and liver tissues (representing two different germ layers) of adult opossums and mice, to understand the global pattern of DNA methylation in metatherians. I compare DNA methylation profiles between male and female samples to investigate roles of DNA methylation in X chromosome inactivation.

3.1 Features of the metatherian genome relevant to the analysis of DNA methylation

I summarise here some features of the opossum genome that are relevant to the analysis of DNA methylation, and compare these to some well-studied vertebrates. The opossum genome is approximately 3.6 Gb, consisting of 8 autosomes (ranging from 260 – 748 Mb), an X chromosome (79 Mb), and unplaced scaffolds (‘chrUn’; 103 Mb; Mikkelsen et al., 2007b). The opossum Y chromosome was not included in the genome assembly. The GC content of the opossum genome is 37.7%, which is lower than eutherian genomes (human GC content 42.5%, mouse CG content 42.2%) and slightly higher than the zebrafish (36.8%).

Opossum CGIs appear similar to eutherian genomes in terms of GC content, with human, mouse and opossum genome having a median CGI CG content of 65-67%, while the median GC content of zebrafish CGIs is 55.3% (Figure 9a). The observed/expected occurrence of CpG sites in opossum CGIs (0.8) is also more similar to human and mouse (0.85 and 0.9, respectively) than to zebrafish (1.17; Figure 9b).

Opossum CGIs are slightly shorter (median length of 451 bp) than human and mouse (median length 569 and 533 bp, respectively), but longer than zebrafish (median length 354 bp). However, opossums are similar to both humans and zebrafish in possessing several very long annotated CGIs (> 10 Kb), while differing noticeably from the mouse, in which the maximum CGI length is approximately 5 Kb (Figure 9c).

CGIs are often found to overlap transcriptional start sites (TSS), but can also be intra- or intergenic (Deaton and Bird, 2011). The relative distribution of CGIs in these genomic categories is most similar between human and opossum, in contrast to the mouse where the majority of CGIs are TSS-associated (Figure 9d). This analysis comes with the caveat that CGIs annotated as intergenic may actually be associated with an as yet unknown gene.

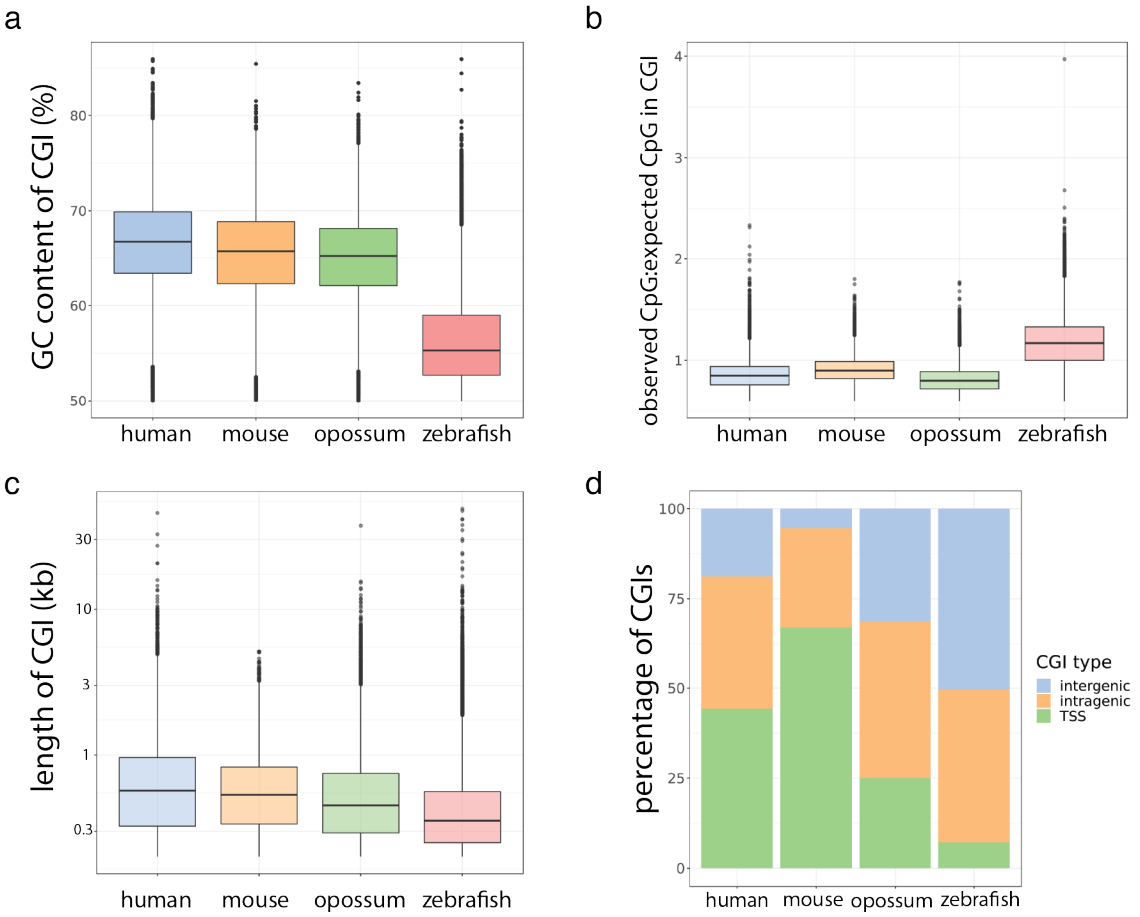


Figure 9. Features of the metatherian genome relevant to the analysis of DNA methylation. **a-d**: Comparison of CpG island (CGI) characteristics between representative vertebrate species. Data derived from UCSC Genome Browser CGI and Ensembl gene annotations. **a**: GC content of CGIs. **b**: ratio of observed: expected CpG content of CGIs. **c**: length of CGIs. **d**: proportion of CGIs overlapping a different genome features; transcriptional start site (TSS, defined as TSS ± 250 bp), intragenic (defined as TSS to transcriptional end site, excluding the first 250 bp), or intergenic (not overlapping TSS or gene).

Sequences homologous to the enzymes involved in DNA methylation in eutherian mammals are also present in the opossum genome (Figure 10). I analysed the expression levels of these enzymes in adult brain and liver of opossum, comparing this data to expression in the same tissues in mouse (Figure 10). This analysis made use of RNA-seq data generated for experiments presented in Chapter 4 (data described in full in Chapter 4.1.2).

DNMT1 and UHRF1 maintain DNA methylation patterns through cell divisions by methylating the daughter strand in newly-replicated hemimethylated DNA at the replication fork (Sharif et al., 2007, Bostick et al., 2007). Both *DNMT1* and *UHRF1* are expressed in brain and liver tissues of the opossum in a manner similar to the mouse homologues (Figure 10a-b).

DNMT3A and DNMT3B are *de novo* methyltransferases (Okano et al., 1999). *DNMT3A* is expressed in opossum brain and liver, while *DNMT3B* is lowly expressed (FPKM <1; Figure 10a). A similar pattern of expression is seen in the mouse homologues (Figure 10b). DNMT3L is a non-catalytic protein partner of the *de novo* methyltransferases (Hata et al., 2002, Suetake et al., 2004). *Dnmt3L* expression is not detectable in adult brain and liver (Figure 10b). Unlike the mouse, opossum *DNMT3L* is expressed in adult liver, though the expression level in brain is very low (FPKM <1; Figure 10a).

Tet proteins are involved in the oxidation of 5mC to 5hmC and further oxidised derivatives, which can precede both passive and active DNA demethylation (Tahiliani et al., 2009, Ito et al., 2011, He et al., 2011). *TET1* and *TET2* are expressed in opossum brain and liver, while *TET3* is very lowly expressed (Figure 10a). In contrast, in the mouse all three *Tet* genes are expressed, though *Tet1* expression was very low in liver (Figure 10b).

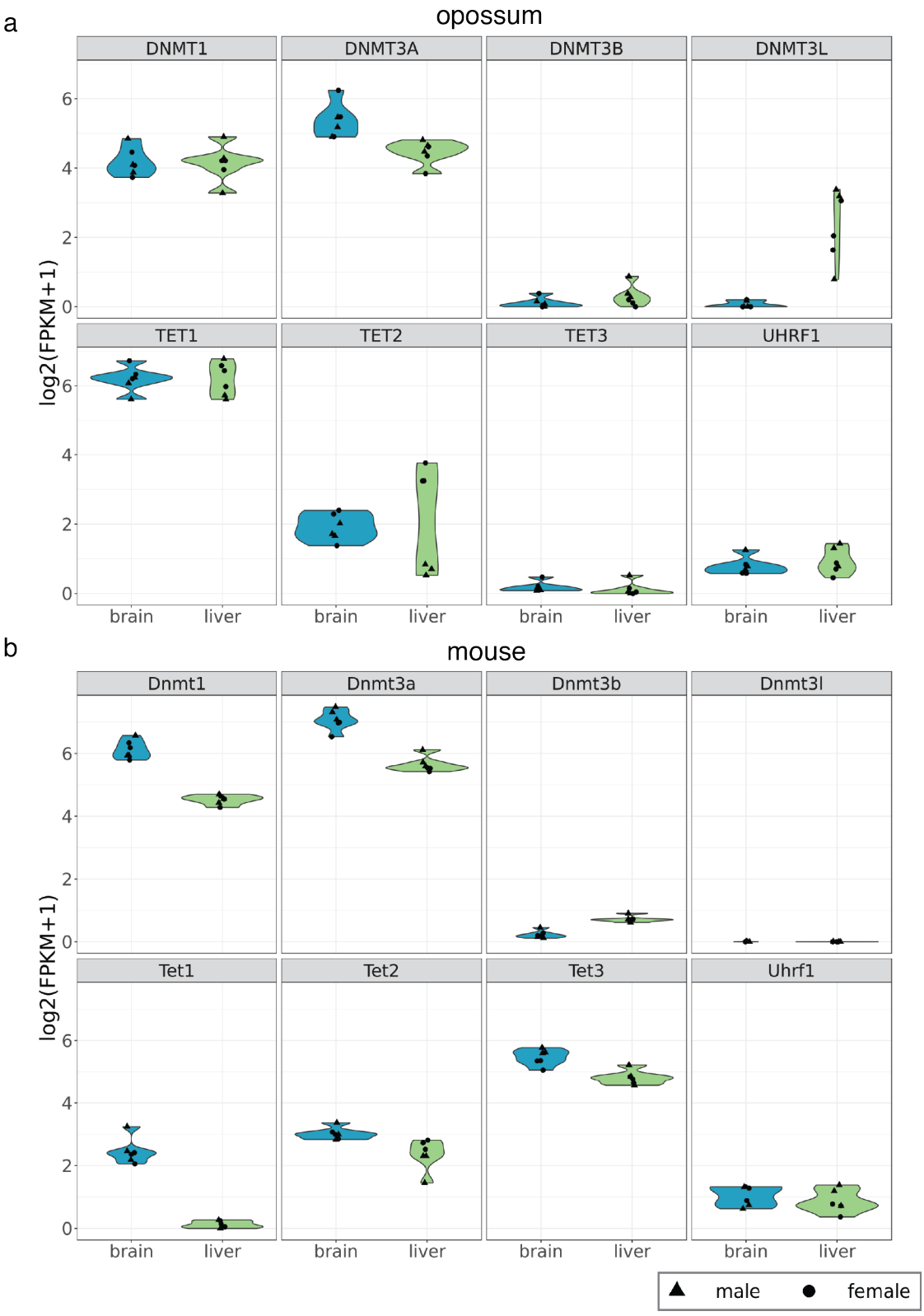


Figure 10: Expression of methylation enzymes. a-b: Violin plots representing the expression levels of key DNA methylation pathway enzymes in opossum and mouse adult tissues. Data comprise three male and three female RNA-seq libraries per condition (represented as overlaid individual points). **a**: Opossum expression levels. **b**: Mouse expression levels.

In summary, the features of opossum genomes related to DNA methylation are largely similar to those of eutherian mammals, with some small differences in overall GC content, and CGI co-occurrence with TSS vs gene bodies. The enzymatic machinery for DNA methylation is present and expressed in a largely similar manner to the mouse. Therefore it might be expected that the global methylation landscape of opossum tissue would generally resemble a typical vertebrate pattern.

3.2 Generation of reduced-representation bisulfite sequencing datasets

RRBS is typically performed using MspI, which creates libraries enriched for CGIs, but digestion with two enzymes can increase coverage of other genomic features (Wang et al., 2013, Lee et al., 2014b, Martinez-Arguelles et al., 2014). As the aims of this project include profiling global methylation in metatherians, achieving coverage of features other than CGIs was desirable. To identify an appropriate enzyme combination for digestion of genomic DNA in RRBS, *in silico* simulations of mouse, opossum, and Tasmanian devil RRBS libraries were generated using several double enzyme combinations (Figure 11).

All double digestions increased total coverage compared to MspI alone (Figure 11). Coverage of CGIs was not notably increased by double digestion, which was expected, as MspI alone captures most CGIs in eutherian genomes (Gu et al., 2011a). Double digestion improved coverage of promoters and gene bodies, though TaqI did not notably enrich for gene bodies. Tsp509I considerably increased coverage of 'other' CpG sites (those not overlapping a CpG island, island shore, promoter, or gene body). Increased coverage of 'other' CpGs is undesirable; methylation at these CpGs sites could be difficult to interpret without genomic context, so Tsp5091 would increase library complexity (and therefore increase the required depth of sequencing) without necessarily returning an increase in informative data.

Based on these results, TaqI and Tsp509I were excluded. BstNI and BfaI were predicted to give similar CpG coverage. BfaI was chosen because it shares buffer compatibility with MspI, permitting simultaneous digestion during library preparation.

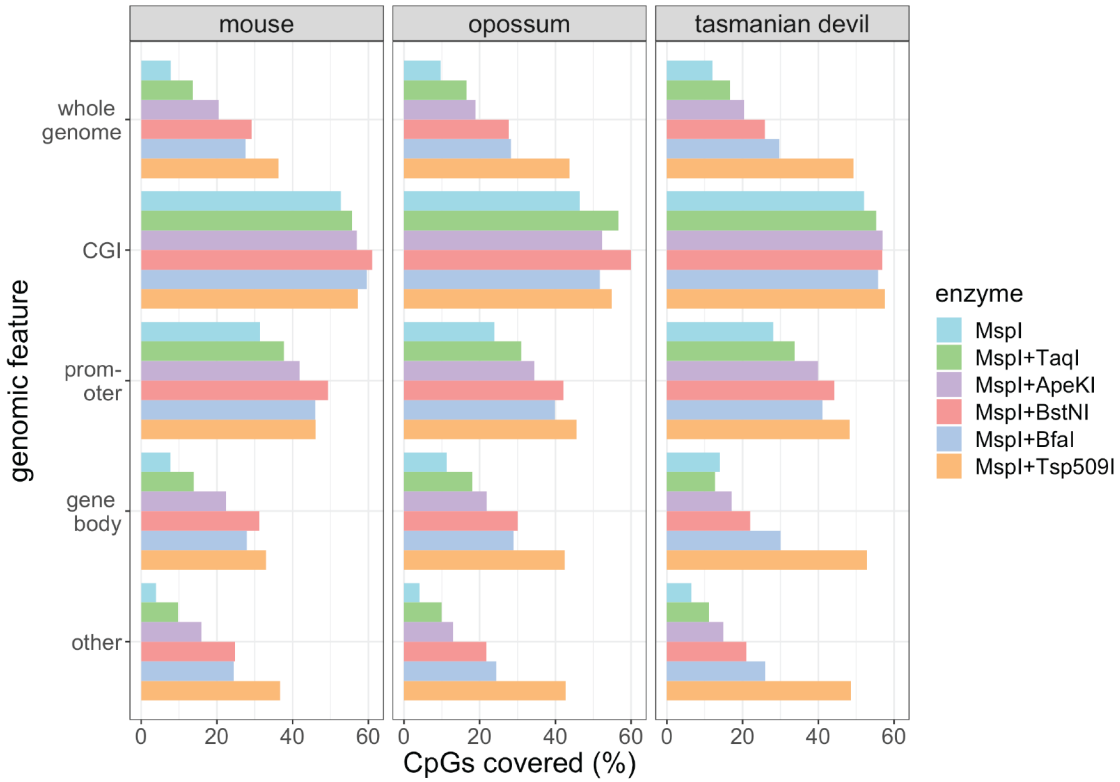


Figure 11. Comparison of restriction enzymes for RRBS. Simulation of CpG coverage in mouse, Tasmanian devil, and opossum RRBS libraries prepared by double digestion with MspI and a second enzyme, size -selected for range 40-400 bp, and sequenced 50 bp paired-end. Genome; coverage of the total CpGs in the genome. CGI; coverage of CpG sites in UCSC-defined CpG islands. Promoter; coverage of promoter CpG sites (defined as the 2 kilobases flanking RefSeq-defined transcriptional start sites). Gene body; coverage of gene body CpG sites (defined as every base from the Refseq- defined transcriptional start sites to transcriptional stop site). Other; any CpG not residing within a CpG island, promoter, or gene body. Simulations were performed by Stuart Horswell (Francis Crick Institute Bioinformatics and Biostatistics Science Technology Platform) using mouse genome build mm10 (05-02-16), Tasmanian devil genome build sarHar1 (03-03-16), and opossum genome build MonDom5 (19-02-16).

RRBS libraries were prepared in biological triplicate from brain and liver of adult male and female opossum, and in biological duplicate from brain and liver of adult male and female mice. High molecular weight genomic DNA (gDNA) was purified from tissues (Figure 12a-b) before being digested with the restriction enzyme combination chosen in the above *in silico* RRBS model. gDNA digested by MspI and BfaI (Figure 12c) appeared as a smear of the size predicted *in silico*, indicating successful digestion (Figure 12d).

The fragment size distributions of prepared RBBS libraries were visualised using an Agilent Bioanalzyer, and were within the expected range for bisulfite libraries (fragments > 150 bp < 500 bp Boyle et al., 2012; Figure 12e-f).

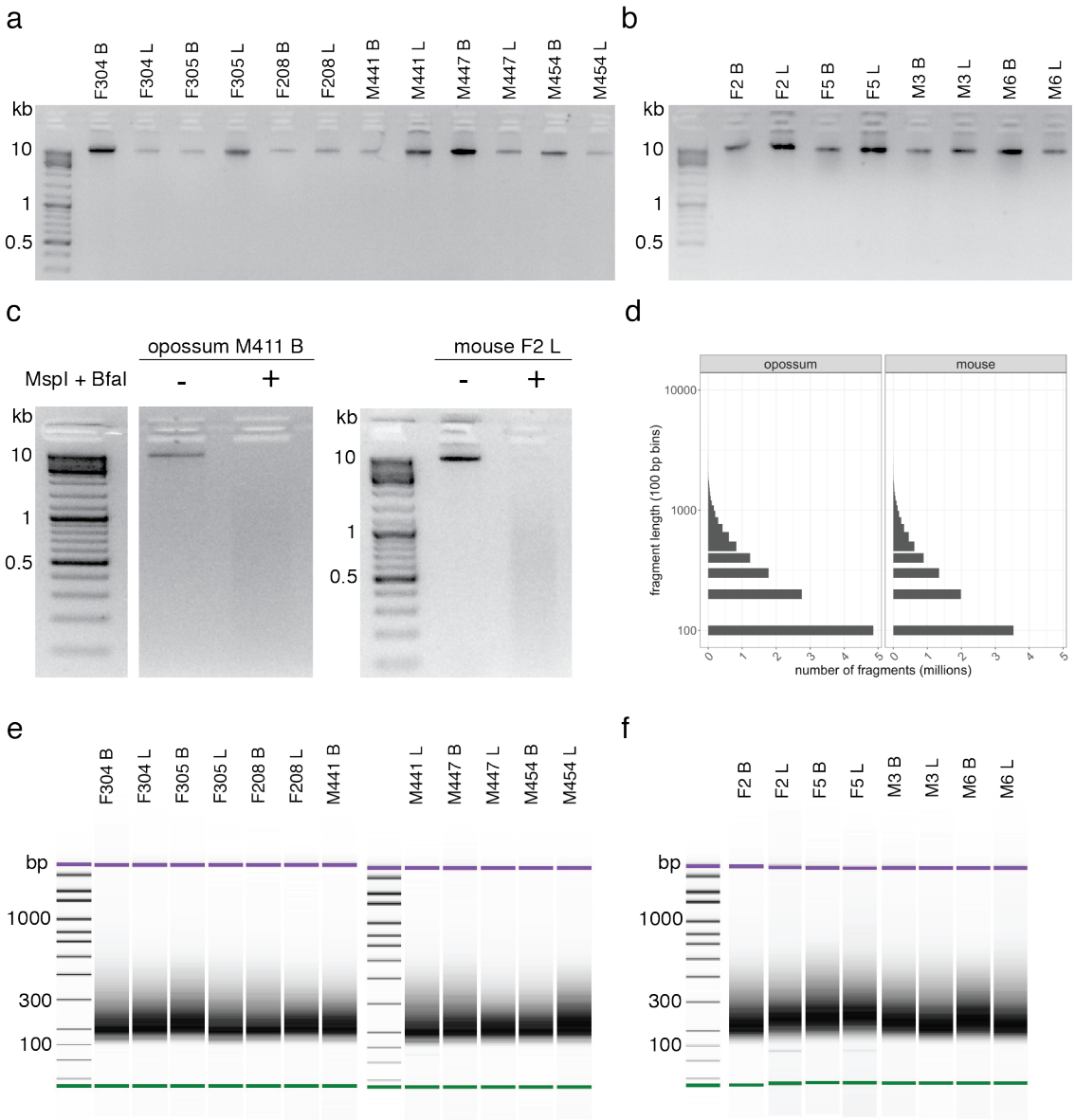


Figure 12. Generation of RRBS libraries. RRBS libraries were generated from the brains and livers of three male and three female adult opossums, and two male and two female adult mice. **a-b**: High molecular weight genomic DNA; **a**: Opossum gDNA, **b**: Mouse gDNA. **c**: gDNA smears following restriction digestion with MspI and BfaI in representative opossum and mouse samples. **d**: Predicted fragment length distribution of gDNA digested with MspI and BfaI in opossum and mouse samples. **e-f**: Bioanalyzer e-gel profiles of resultant RRBS libraries. Bioanalyzer assays were performed by Deb Jackson (Francis Crick Institute Advanced Sequencing Facility). **e**: Opossum libraries, **f**: Mouse libraries.

Library	Species	Total reads (millions)	Bases retained after filtering (%)	Mapping efficiency (%)	C methylated in CHH context (%)	C methylated in CpG context (%)
F304 Brain	Opossum	97.4	90	70	3	60.5
F304 Liver	Opossum	106.1	93.5	71	1	56
F305 Brain	Opossum	91.5	95.3	71.5	2	61
F305 Liver	Opossum	96.5	90.2	74	1	52.7
F208 Brain	Opossum	80.8	90.2	74.7	2	61.3
F208 Liver	Opossum	98.2	95.5	71.3	2	55
M441 Brain	Opossum	70.4	91	73.7	2	60.3
M441 Liver	Opossum	139.4	91.6	73	2	51
M447 Brain	Opossum	77.9	92.2	74.3	2	62.7
M447 Liver	Opossum	95.7	92.7	73.5	2	46
M454 Brain	Opossum	121.9	91	73.7	2	64.3
M454 Liver	Opossum	124.7	90.2	73.3	1	61.7
F2 Brain	Mouse	135.2	90.3	72.7	2.7	62
F2 Liver	Mouse	78.5	90.8	76.3	2	58.3
M5 Brain	Mouse	94.2	93	72	2	52
M5 Liver	Mouse	48.5	91.8	75	2	56.5
F3 Brain	Mouse	103.9	92	73.3	2	61.3
F3 Liver	Mouse	84.5	90.4	72.5	1.5	55
M6 Brain	Mouse	119.1	93	69.3	2	50
M6 Liver	Mouse	114.9	90.8	72.3	2	59.3

Table 3. RRBS library statistics

100 bp paired-end sequencing of opossum and mouse RRBS libraries was performed over several replicate runs in order to achieve a minimum of 50 million reads per library, culminating in between 50 -140 million reads per library (Table 3). 90 - 95% of bases were retained following adapter and quality trimming, and libraries mapped with efficiencies between 69-76% (Table 3).

A concern when interpreting bisulfite sequencing libraries is whether or not the bisulfite conversion has reached completion. In vertebrate genomes, where methylation in a CHH context is rare, it is expected that bisulfite conversion of CHH cytosines will be essentially complete (i.e. 0% methylation). The CHH methylation rate can thus be used as a proxy for bisulfite non-conversion rate. The average CHH methylation rate in these libraries was 1.9%, implying a bisulfite non-conversion rate of approximately 2% (Table 3). The average methylation rate of cytosines in a CpG context was 57.2% (Table 3).

At initial sequencing, trimming, and mapping, CpG sites were covered at an average depth of 3.6x (Figure 13a-b). After resequencing and merging of replicate libraries from the different sequencing runs, the average coverage depth of captured CpGs increased to 4.6x. The *in silico* RRBS model (Stuart Horswell, Crick Bioinformatics and Biostatistics) predicted that MspI + BfaI RRBS libraries would capture approximately 28% of genomic CpGs. In the initial run, many libraries did not reach the predicted breadth of coverage. Following resequencing, the breadth of coverage increased, reaching or exceeding 27% in all libraries (Figure 13a-b).

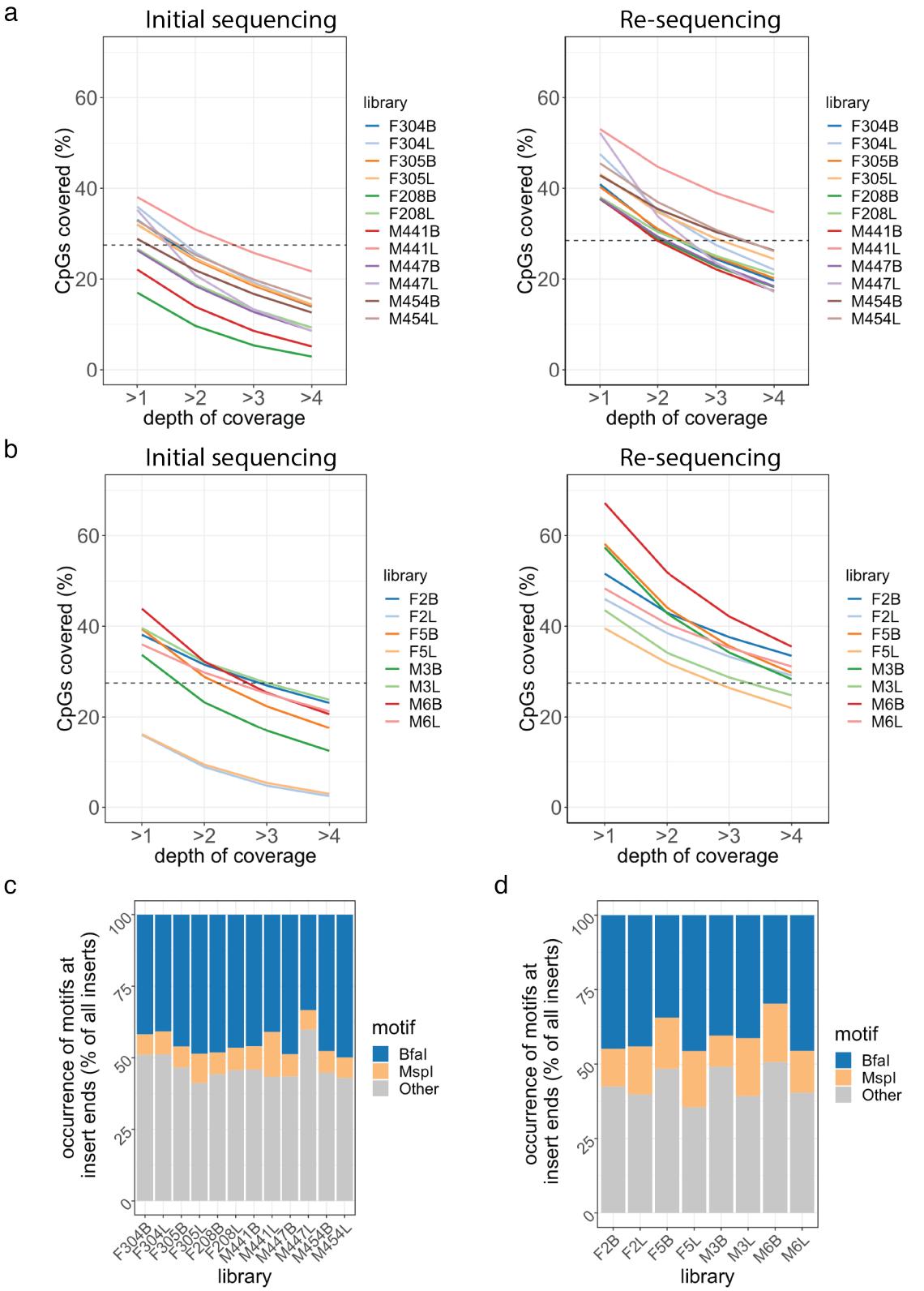


Figure 13. Coverage and restriction enzyme motif analysis of RRBS libraries. a-b: Percentage of CpG sites covered at different depths in RRBS libraries. Horizontal dotted line represents coverage of genomic CpGs predicted by an *in silico* RRBS model. **a**: Opossum. **b**: Mouse. **c-d**: Percentage of read ends containing an MspI or BfaI digestion motif. **c**: Opossum. **d**: Mouse.

As RRBS libraries are prepared using restriction digestion of gDNA, it would be expected that read ends contain the restriction site motif of the enzymes used during library preparation. The three base pairs beginning and ending each insert were searched for motifs representing the MspI and BfaI enzyme restriction sites (Figure 13c-d). While MspI and BfaI motifs were present, approximately half of all insert ends were not derived from either restriction enzyme, implying that inserts derived from random fragmentation of gDNA contributed significantly to these libraries. However, as the aim of using a double-enzyme digest was to increase the breadth of genome coverage, capture of some random genome fragments did not invalidate the usefulness of these libraries.

A final dataset for analysis was prepared by pooling samples according to condition (i.e. sex, tissue, and species). For retention of a CpG site in the resultant dataset, the following criterion was employed: a CpG must be covered by a minimum of 4 reads in at least 2 samples per condition. This resulted in a final set of 2,416,540 CpGs in opossum, and 2,870,372 CpGs in mouse, with a minimum coverage of 8 reads per CpG. I subsequently analysed global methylation patterns in this RRBS dataset.

3.3 The *M. domestica* global DNA methylation landscape resembles a typical vertebrate pattern

Examination of global methylation features in opossum and mouse RRBS data revealed that the metatherian genome displays a pattern of DNA methylation typical of vertebrate genomes. Like the mouse, the opossum genome displays a bimodal methylation pattern when considering the bulk genome (100 bp tiles, Figure 14a-b). In both species there is no apparent difference in methylation between tissues or between the sexes.

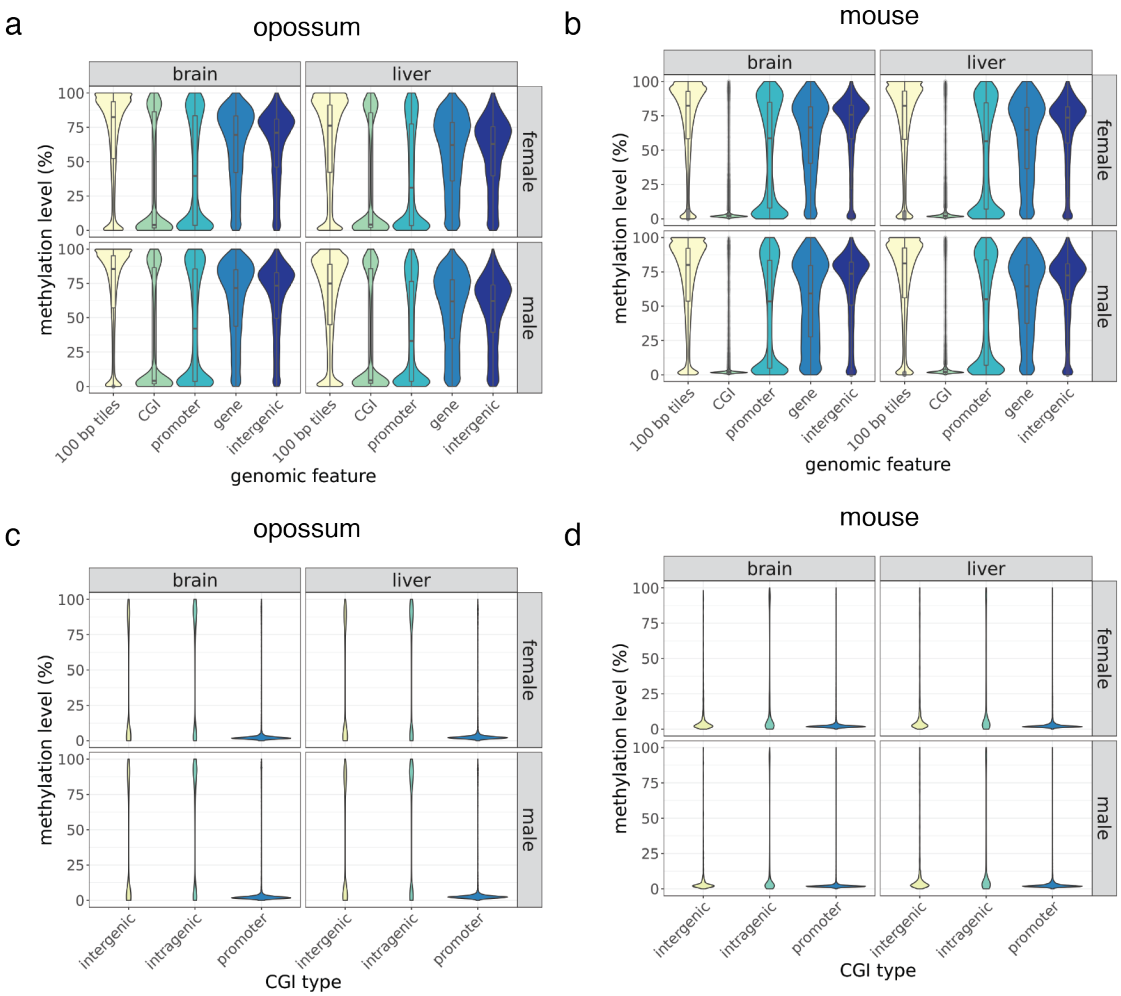


Figure 14. Global methylation patterns of opossum and mouse genomes. Violin plots depicting the methylation level of genomic features in opossum and mouse genomes. Methylation level is the percentage of methylated CpGs/covered CpGs within a region. Opossum data are the pool of biological triplicates, mouse data are the pool of biological duplicates. CpGs were included if covered by a minimum of 4 reads in at least 2 replicates per condition. **a-b**:100 bp tiles; non-overlapping 100 bp tiles; CGI; UCSC defined CpG islands; Promoter; Ensembl defined TSS – 2kb/+ 200 bp; Gene; Ensembl defined genes from TSS to transcription end site; Intergenic; all genomic regions not included in 'gene' category. **a**: opossum, n_{tiles} = 1387252, n_{CGI} = 19978, $n_{promoter}$= 18506, n_{gene} = 22570, $n_{intergenic}$ = 23823. **b**: mouse, n_{tiles} = 1574508, n_{CGI} = 14833, $n_{promoter}$ = 38179, n_{gene} = 35109, $n_{intergenic}$ = 31752. **c-d**: Methylation level of CGIs in different genomic contexts in opossum and mouse genomes. CGI type defined as overlap with regions defined as promoter, gene, or intergenic as above. **c**: opossum, $n_{intergenic}$ = 5265, $n_{intragenic}$ = 8548, $n_{promoter}$ = 6165. **d**: mouse, $n_{intergenic}$ = 565, $n_{intragenic}$ = 3641, $n_{promoter}$ = 10627.

The origin of the bimodal distribution is evident when examining methylation level across different genomic features. As is typical for a vertebrate genome, genes and intergenic regions are hypermethylated, while CGIs and promoters are hypomethylated (Figure 14a-b). Unlike the mouse, in opossum some CGIs are hypermethylated. While CGIs are often

found at promoters, some CGIs are intragenic or intergenic (Illingworth and Bird, 2009). Examination of DNA methylation at CGIs in different genomic contexts reveals a similar trend in opossum and mouse, where promoter CGIs are the most hypomethylated, but many inter- and intra-genic CGIs are also hypomethylated (Figure 14c-d).

Having established that, as expected, the metatherian global DNA methylation displays typical vertebrate characteristics, I went on to investigate the pattern of methylation on the X chromosome, relating this to XCI by comparing the X chromosome to the autosomes, and by comparing the X chromosomes between males and females.

3.4 DNA methylation landscape of the eutherian X chromosome

Firstly, I examined X-chromosome methylation patterns in male and female mice. This served two purposes; the validation of the technique by replicating previously reported methylation patterns of the mouse X chromosome, and the generation of a methylation profile of eutherian X methylation for direct comparison with the opossum data generated using an identical technique.

The methylation level at X-linked loci in male samples is representative of the single, active X chromosome. In females, the observed methylation level derives from the average signal from two X chromosomes, one active and one inactive. Therefore, a difference in methylation between the Xa and the Xi would be observed as an intermediate methylation level in female samples when compared to males.

The methylation of the mouse X chromosome is similar in distribution to the methylation of the autosomes. In both males and females, in both tissues profiled, the majority of 100 bp tiles have a high methylation level (blue violin plots, Figure 15a). However, averaging over all 100 bp tiles, the male X chromosome has a slightly higher methylation level than the female X chromosome in both brain (median male methylation: 85%, median female methylation: 78.6%, $p < 2.2e\text{-}16$, Mann-Whitney test) and liver (median male methylation: 83.3%, median female methylation: 78.6%, $p < 2.2e\text{-}16$, Mann-Whitney test). The average methylation level of the male X chromosome is also slightly higher

than that of autosomes in brain (median autosome methylation: 81.03%, $p < 2.2e-16$, Mann-Whitney test) and liver (median autosome methylation: 80%, $p < 2.2e-16$, Mann-Whitney test). This suggests that there is an X chromosome-specific phenomenon governing methylation levels.

Confirming the observation made using 100 bp tiles, when differential methylation of individual CpG sites was calculated between females and males (sites with q-value <0.01 and a methylation difference ≥ 25%, Fisher's exact test), differentially-methylated sites occurred more frequently on the X chromosome than on autosomes. In both brain and liver, fewer than 1% of covered CpG sites on autosomes were differentially methylated between females and males (Figure 15b). A larger fraction of covered CpG sites were differentially methylated between males and females on the X chromosome (Figure 15b). In brain, 5.5% of covered CpG sites were hypermethylated in females, and 1.7% were hypomethylated. In liver, 1.2% of covered CpG sites were hypermethylated in females, and 1.1% were hypomethylated.

Given the observation that the male X is hypermethylated relative to female (Figure 15a), it was surprising that the majority of differentially-methylated CpG sites were hypermethylated in female (Figure 15b). This could be due to the methylation difference threshold of ≥ 25% applied in the differential methylation testing. As the overall difference in methylation between male and female X chromosomes is in the range of 10%, many regions may have been excluded. It therefore interesting to consider at which features of the X chromosome the differences between male and female originate.

In brain, 72.4% of X-linked hypermethylated CpG sites were in genes, and 19.8% were in CGIs, while 47.6% of hypomethylated sites were in genes, and 2.1% were in CGIs. In liver, 66.8% of hypermethylated sites were in genes, and 21.2% were in CGIs, while 47.6% of hypomethylated sites were in genes, and 2% were in CGIs. These results suggest that gene bodies can be either hyper- or hypomethylated in females relative to males, while CGIs are subject to female-specific hypermethylation.

By examining methylation levels at different genomic features, exactly these two patterns can be clearly observed. Female X-chromosome CGI methylation (median female X methylation; brain: 23.9%, liver: 21.1%) is elevated relative to autosomes (median female autosome methylation; brain: 2.1%, liver: 2.2%) and the male X (median male X methylation; brain: 1.8%, liver: 2.3%; orange violin plots, Figure 15a).

Female X-chromosome gene body methylation displays a shift towards the middle of the distribution (both fewer lowly methylated and fewer highly methylated sites) relative to autosomes and the male X (green violin plots, Figure 15a). A similar shift can be observed in both promoter (purple violin plots, Figure 15a), and intergenic regions (red violin plots, Figure 15a), implying that the entire X chromosome may bear differences in methylation level.

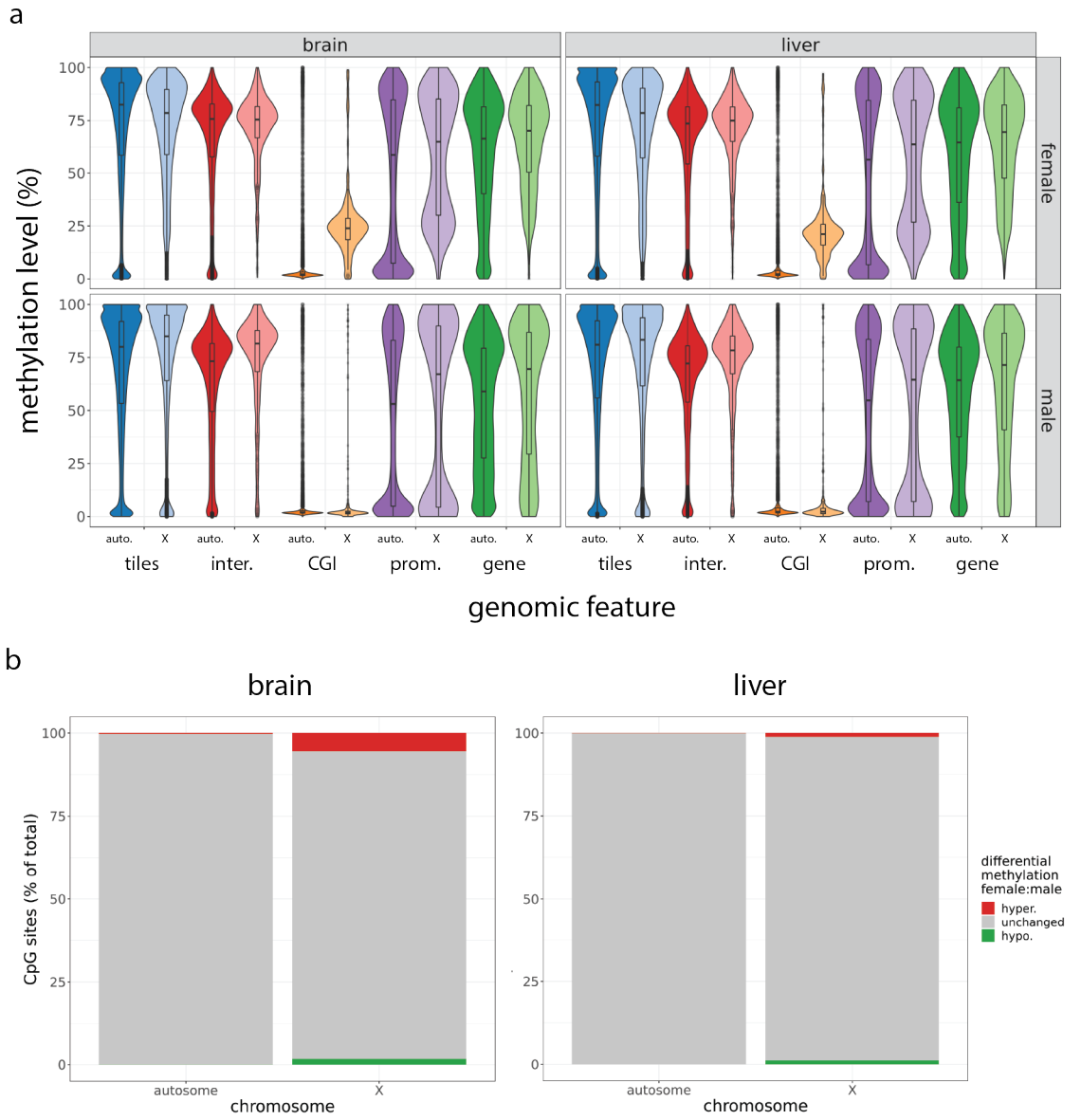


Figure 15. DNA methylation pattern of the eutherian X chromosome. DNA methylation of the mouse X chromosome compared to autosomes. Data are the pool of biological duplicates. Methylation level is the percentage of methylated CpGs/ covered CpGs within a region. CpGs were included if covered by a minimum of 4 reads in at least 2 replicates per condition. **a**: Violin plots depicting the methylation level of genomic features on the grouped autosomes and the X chromosome. 100 bp tiles; non-overlapping 100 bp tiles; CGI; UCSC defined CpG islands; Promoter; Ensembl defined TSS – 2kb/+ 200 bp; Gene; Ensembl defined genes from TSS to transcription end site; Intergenic; all genomic regions not included in 'gene' category. $n_{AutoTiles} = 1539517$, $n_{XTiles} = 34287$, $n_{AutoIntergenic} = 29994$, $n_{XIntergenic} = 1697$, $n_{AutoCGI} = 14449$, $n_{XCGI} = 382$, $n_{AutoPromoter} = 37016$, $n_{XPromoter} = 1162$, $n_{AutoGene} = 33989$, $n_{XGene} = 1116$. **b**: Percentage of CpG sites found to be differentially methylated between female and male chromosomes (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test). Red: hypermethylated in female, green: hypomethylated in female, grey: unchanged between the sexes.

Assuming that the Xa in females resembles the single active X in males, these results imply that the Xi is hypomethylated relative to the Xa, with the exception of CGI regions, where the Xi is hypermethylated. These data replicate the well-reported phenomenon of Xi CGI hypermethylation (Gendrel et al., 2012), and align with previous reports of Xa intragenic and intergenic hypermethylation (Hellman and Chess, 2007). However, the scale of the difference between Xa and Xi at regions outside CGIs is very small, while at CGIs the difference is approximately 20%.

The examination of X-chromosome CGIs and gene bodies at an individual locus level (Figure 16a-b) confirms that the averaged methylation distributions described above hold true across the majority of individual loci. With the exception of several CGIs that are highly methylated in both sexes, individual CGIs display the described pattern of negligible methylation in males, and approximately 25% methylation in females (Figure 16a). The majority of gene bodies display methylation levels of approximately 80% in males and females (Figure 16b). The locus surrounding the X-linked *Hprt* gene is presented as an example of the methylation profile across a single gene, with lower methylation over CGIs, and high methylation over the gene body and the intergenic regions (Figure 16c).

Having generated a mouse RRBS dataset that replicated previous findings, thereby validating my technical approach, I went on to analyse metatherian methylation in the opossum dataset.

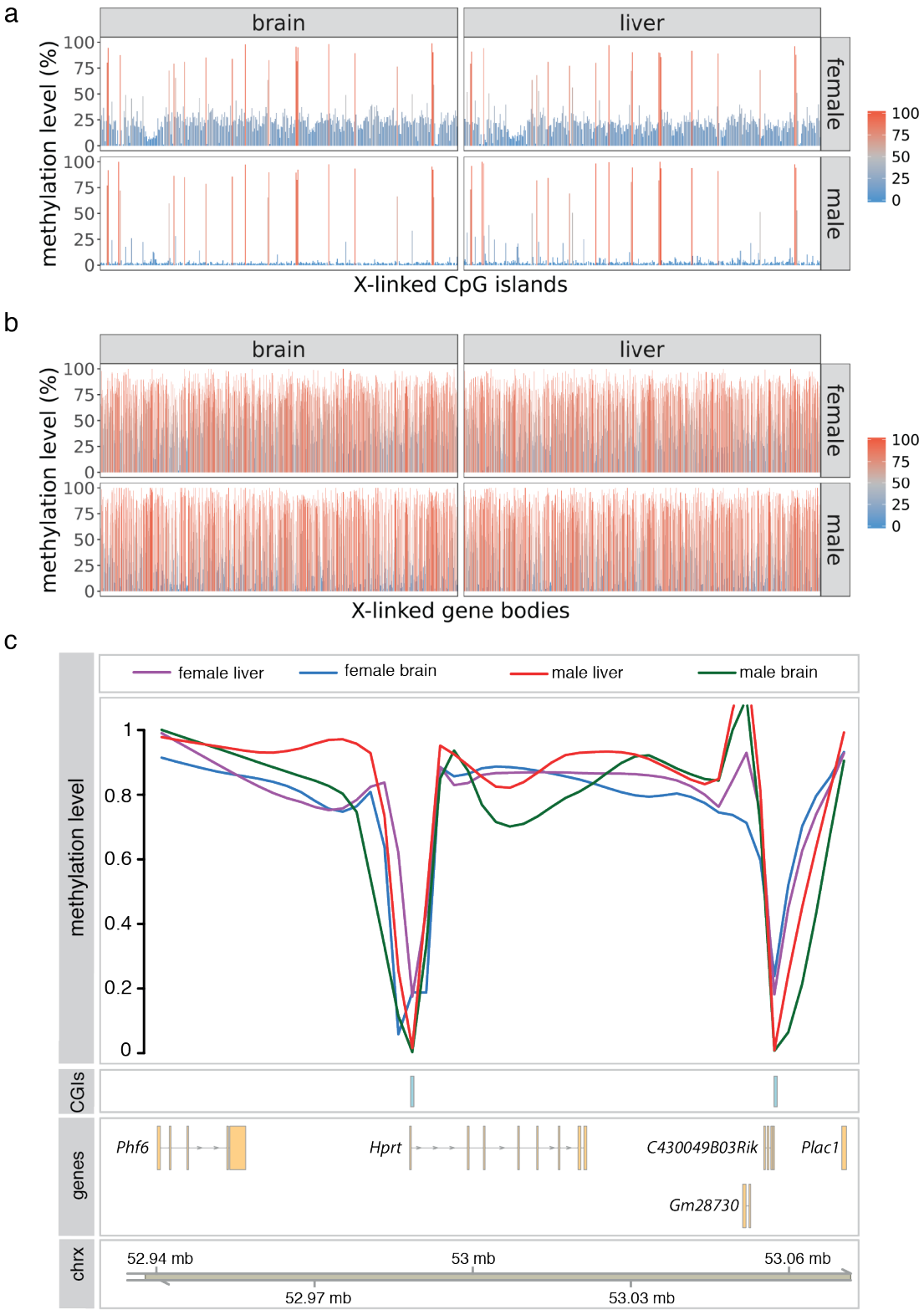

a
brain
liver
female
male
methylation level (%)
100
75
50
25
0
X-linked CpG islands
b
brain
liver
female
male
methylation level (%)
X-linked gene bodies
c
female liver
female brain
male liver
male brain
methylation level
1
0.8
0.6
0.4
0.2
0
CGIs
genes
Phf6
Hprt
C430049B03Rik
Plac1
Gm28730
chrx
52.94 mb
53 mb
53.06 mb
52.97 mb
53.03 mb

Figure 16. Locus specific CGI and gene body DNA methylation patterns of the eutherian X chromosome. a-b: Bar plots depicting methylation level of individual genomic features on the mouse X chromosome. Methylation level is the percentage of methylated CpGs/covered CpGs within a region. **a**: CGIs (UCSC defined CpG islands; n = 382), ordered along the X axis according to position on the X chromosome. **b**: Gene bodies (Ensembl defined genes from TSS to transcription end site; n = 1116), ordered along the X axis according to position on the X chromosome. **c**: Genomic visualisation plot of the *Hprt* locus. Methylation is plotted as methylation level per CpG site (n = 127), smoothed by the 'loess' function of the R package 'Gviz'. Data are the pool of biological duplicates. CpGs were included if covered by a minimum of 4 reads in at least 2 replicates per condition.

3.5 DNA methylation landscape of the metatherian X chromosome

Initially I examined chromosome-wide distributions of methylation, using 100 bp non-overlapping tiles (blue violin plots, Figure 17a). The male X chromosome and autosomes were similarly hypermethylated, though the X chromosome had slightly higher methylation in both tissues profiled ($p < 2.2e-16$, Mann-Whitney test). Most 100 bp tiles displayed methylation greater than 70% in both brain (median male X methylation: 88.6%, median male autosome methylation: 85.5%) and liver (median male X methylation: 79.7%, median male autosome methylation: 74.8%). The methylation distribution of the female X chromosome differed dramatically to both the autosomes, and the male X in both tissues profiled ($p < 2.2e-16$, Mann-Whitney test). 100 bp tiles on the X chromosome in females had a median methylation level of 42.9% in brain and 41.9% in liver, while 100 bp tiles on autosomes had a median methylation level of 83.3% in brain and 77.2% in liver (blue violin plots, Figure 17a).

Differential methylation of individual CpG sites was calculated between females and males (sites with q-value <0.01 and a methylation difference ≥ 25%, Fisher's exact test; Figure 17b). The vast majority of differentially-methylated sites were on the X chromosome, with fewer than 2% of covered CpG sites on autosome showing differential methylation. In comparison, on the X chromosome, in brain, 0.2% of CpG sites were hypermethylated in females, and 37.3% of sites were hypomethylated. 42.1% of these hypomethylated sites were in genes, and 3% were in CGIs. In liver, 0.2% of CpG sites were hypermethylated in females, and 32.9% of sites were hypomethylated. 47.6% of these hypomethylated sites were in genes, and 6.1% were in CGIs. Therefore the X

chromosome in females is dramatically hypomethylated relative to the male, and the sites of differential methylation are largely found outside of CGIs.

This observation is confirmed by examining methylation distributions as a function of different genomic features (Figure 17a). CGI methylation is uniformly low, with female CGI methylation between 3-4% for X and autosomes, and male CGI methylation between 2-4% for X and autosomes (Figure 17a, orange violin plots). In contrast, gene bodies display approximately half the methylation level in females (median brain methylation: 40.3%, median liver methylation: 39.1%) compared to males (median brain methylation: 73.2%, median liver methylation: 68.3%; Figure 17a, green violin plots). The same pattern of intermediate methylation level in females is apparent at intergenic regions (median female brain methylation: 39.9%, median female liver methylation: 38.7%, median male brain methylation: 76%, median male liver methylation: 68.4%; Figure 17a, red violin plots). Promoter regions also show hypomethylation on the X chromosome in females, relative to both the male X and to autosomes (Figure 17a, purple violin plots).

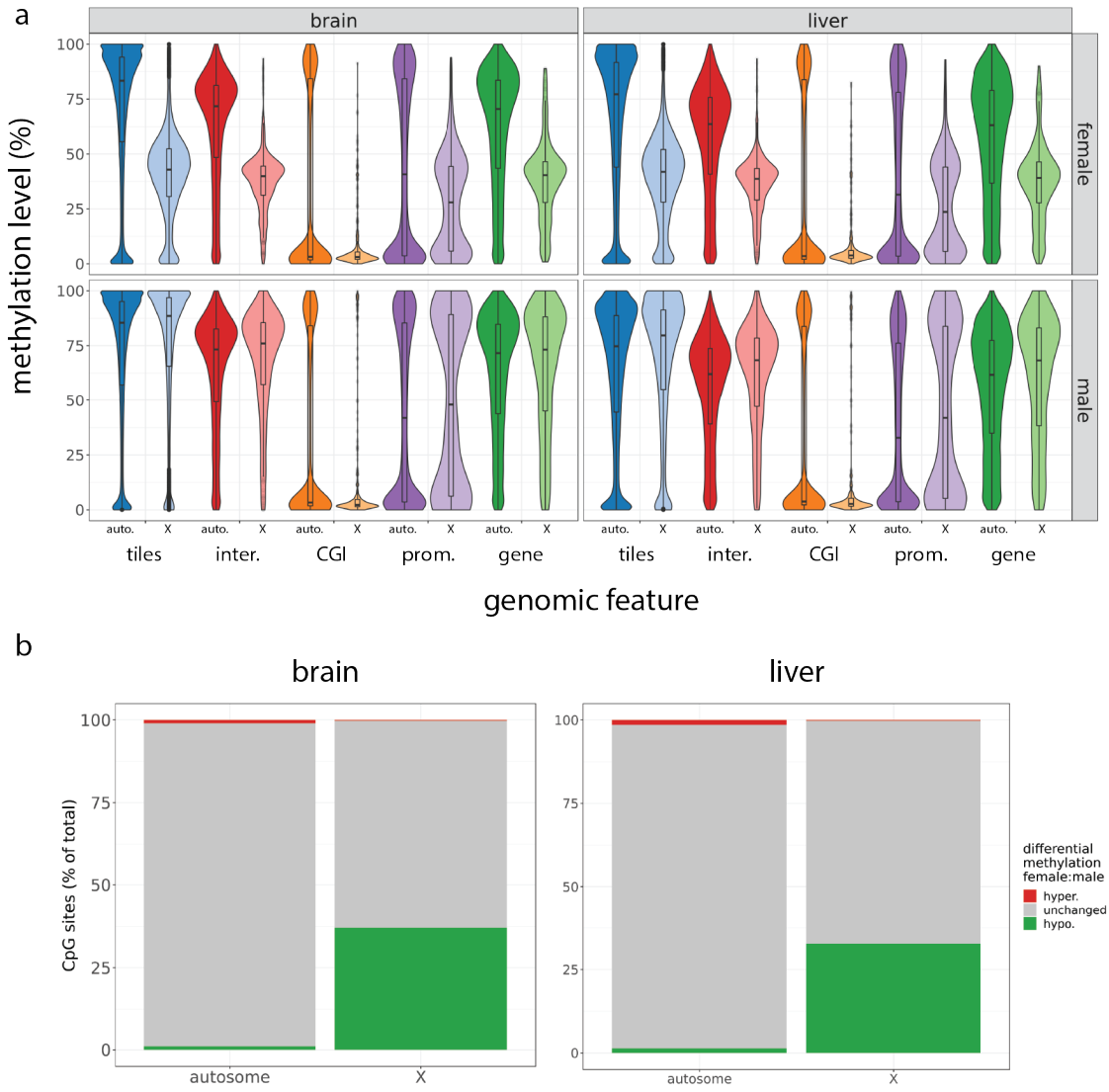


Figure 17. DNA methylation patterns of the metatherian X chromosome. DNA methylation of the opossum X chromosome compared to autosomes. Data are the pool of biological triplicates. Methylation level is the percentage of methylated CpGs/ covered CpGs within a region. CpGs were included if covered by a minimum of 4 reads in at least 2 replicates per condition. **a**: Violin plots depicting the methylation level of genomic features on the grouped autosomes and the X chromosome. 100 bp tiles; non-overlapping 100 bp tiles; CGI; UCSC defined CpG islands; Promoter; Ensembl defined TSS – 2kb/+ 200 bp; Gene; Ensembl defined genes from TSS to transcription end site; Intergenic; all genomic regions not included in 'gene' category. $n_{AutoTiles} = 1294462$, $n_{XTiles} = 36525$, $n_{AutoIntergenic} = 23253$, $n_{XIntergenic} = 569$, $n_{AutoCGI} = 17419$, $n_{XCGI} = 392$, $n_{AutoPromoter} = 17992$, $n_{XPromoter} = 514$, $n_{AutoGene} = 21992$, $n_{XGene} = 578$. **b**: Percentage of CpG sites found to be differentially methylated between female and male chromosomes (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test). Red: hypermethylated in female, green: hypomethylated in female, grey: unchanged between the sexes.

With the assumption that the Xa in females resembles the single active X in males, these results imply that the Xi is vastly hypomethylated relative to the Xa. This data also confirms previous reports that the metatherian Xi does not display CGI hypermethylation (Kaslow and Migeon, 1987, Loebel and Johnston, 1996, Wang et al., 2014b, Waters et al., 2018).

The examination of X-chromosome CGIs and gene bodies at an individual locus level (Figure 18a-b) confirms that the averaged methylation distributions described above hold true across the majority of individual loci. With the exception of several CGIs that are highly methylated in males, and intermediately methylated in females, individual CGIs display the described pattern of negligible methylation in males and females (Figure 18a). The majority of gene bodies display methylation levels of approximately 70% in males and approximately 40% in females (Figure 18b). Intriguingly, a number of opossum X-linked genes did not conform to the general trend of intermediate methylation in females; rather these genes had methylation levels as high as 70%; similar levels to the equivalent male locus (Figure 18b).

The *HPRT1/RSX* locus is presented as an example of the methylation profile across single genes, showing lower methylation over CGIs, and higher methylation over the gene body and the intergenic regions (Figure 18c). Differential methylation between males and females can be seen at the *HPRT1* and *RSX* gene body.

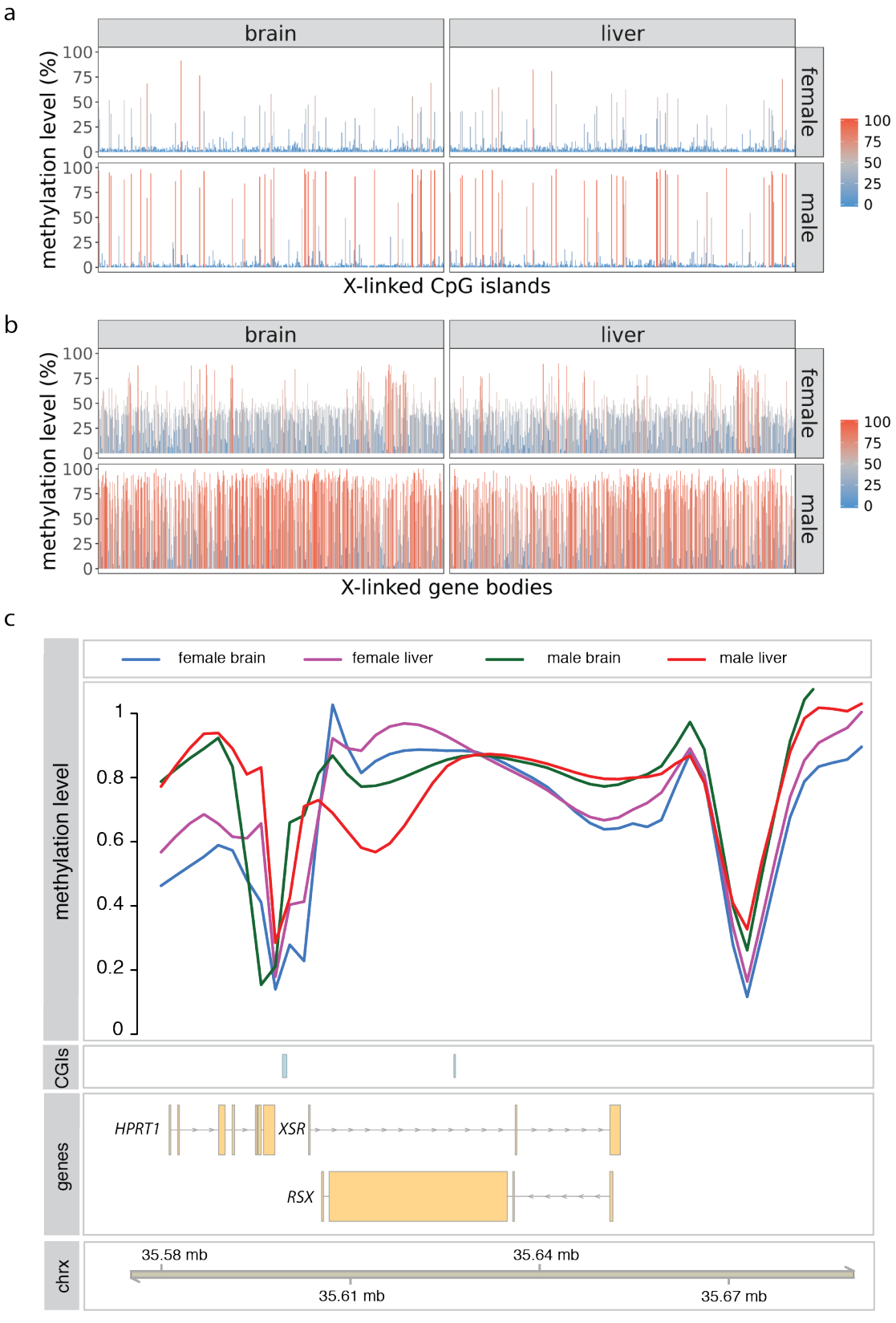

a
brain
liver
female
male
methylation level (%)
100
75
50
25
0
X-linked CpG islands
b
brain
liver
female
male
methylation level (%)
X-linked gene bodies
c
female brain
female liver
male brain
male liver
methylation level
1
0.8
0.6
0.4
0.2
0
CGIs
genes
HPRT1
XSR
RSX
chrx
35.58 mb
35.64 mb
35.61 mb
35.67 mb

Figure 18. Locus specific CGI and gene body DNA methylation patterns of the metatherian X chromosome. a-b: Bar plots depicting methylation level of individual genomic features on the opossum X chromosome. Methylation level is the percentage of methylated CpGs/covered CpGs within a region. **a**: CGIs (UCSC defined CpG islands; n =392), ordered along the X axis according to position on the X chromosome. **b**: Gene bodies (Ensembl defined genes from TSS to transcription end site; n = 578), ordered along the X axis according to position on the X chromosome. **c**: Genomic visualisation plot of the *HPRT1/RSX* locus. Methylation is plotted as methylation level per CpG site (n = 88), smoothed by the 'loess' function of the R package 'Gviz'. Data are the pool of biological triplicates. CpGs were included if covered by a minimum of 4 reads in at least 2 replicates per condition.

3.6 Methylation patterns of metatherian escape genes

I hypothesised that those loci where female and male methylation levels were equivalent may represent XCI-escape genes, where it might be expected that the 'active' hypermethylation state exists on both the Xa and Xi allele in females, resulting in an average level equivalent to the single Xa allele in males. Any locus with a female methylation level approaching the male level could therefore be considered a putative escape gene.

I calculated the female:male methylation ratio of each gene body (excluding those genes where both sexes had <10% methylation) and overlapped this data with a list of opossum genes previously reported to escape XCI based on analysis of allele-specific expression in foetal brain and extra-embryonic membranes (Wang et al., 2014b; Figure 19a). Of the 24 escape genes previously identified, 22 are covered in this data (red bars, Figure 19a). The 22 known escape genes have an average female:male methylation ratio of 1 (average brain ratio =1.01, sd = 0.19; average liver ratio=1.03, sd = 0.18). The average female:male ratio of non-escape/escape status unknown genes was 0.6 (average brain ratio = 0.64, sd = 0.21, average liver ratio = 0.54, sd = 0.19; Figure 19a). These results strongly imply that metatherian XCI escape genes have a unique gene body methylation pattern, and suggest that it may be possible to predict novel escape genes based on a female:male methylation comparison.

At genes that escape XCI, the Xi-linked allele often expresses at a lower level than the Xa-allele (Berletch et al., 2015). If levels of methylation and transcription are linked, it

might therefore be expected for the Xi-allele to have a lower methylation level than the Xa-allele. I therefore sought to identify a cut-off methylation level that encompassed the majority of annotated XCI-escape genes. I chose the threshold of female:male ratio > 0.8; this represents the average methylation ratio of annotated XCI-escape genes (~1) minus the standard deviation (~0.2) and should therefore encompass most known escape genes.

In brain, this cut-off encompassed 19 of the 22 annotated escape genes. Two of the excluded genes, *HMGB3* and *FRMD7,* had ratios of 0.78, i.e. very close to the 0.8 cutoff. In the liver, 21/22 annotated escape genes have a female:male ratio greater than 0.8. The gene excluded in liver, *ALAS2*, had a ratio equivalent to the non-escape/ escape status unknown classes in both tissues (brain ratio: 0.54, liver ratio: 0.59). It is therefore likely that *ALAS2* escapes XCI in the tissues profiled in Wang et al. (2014b), but not in adult brain or liver. Based on the threshold of female:male ratio of 0.8, a further 70 genes in brain, and 85 genes in liver might be considered putative escape genes (Appendix 7.2) The XCI escape status of these genes could be confirmed by investigating their expression in an allele-specific assay.

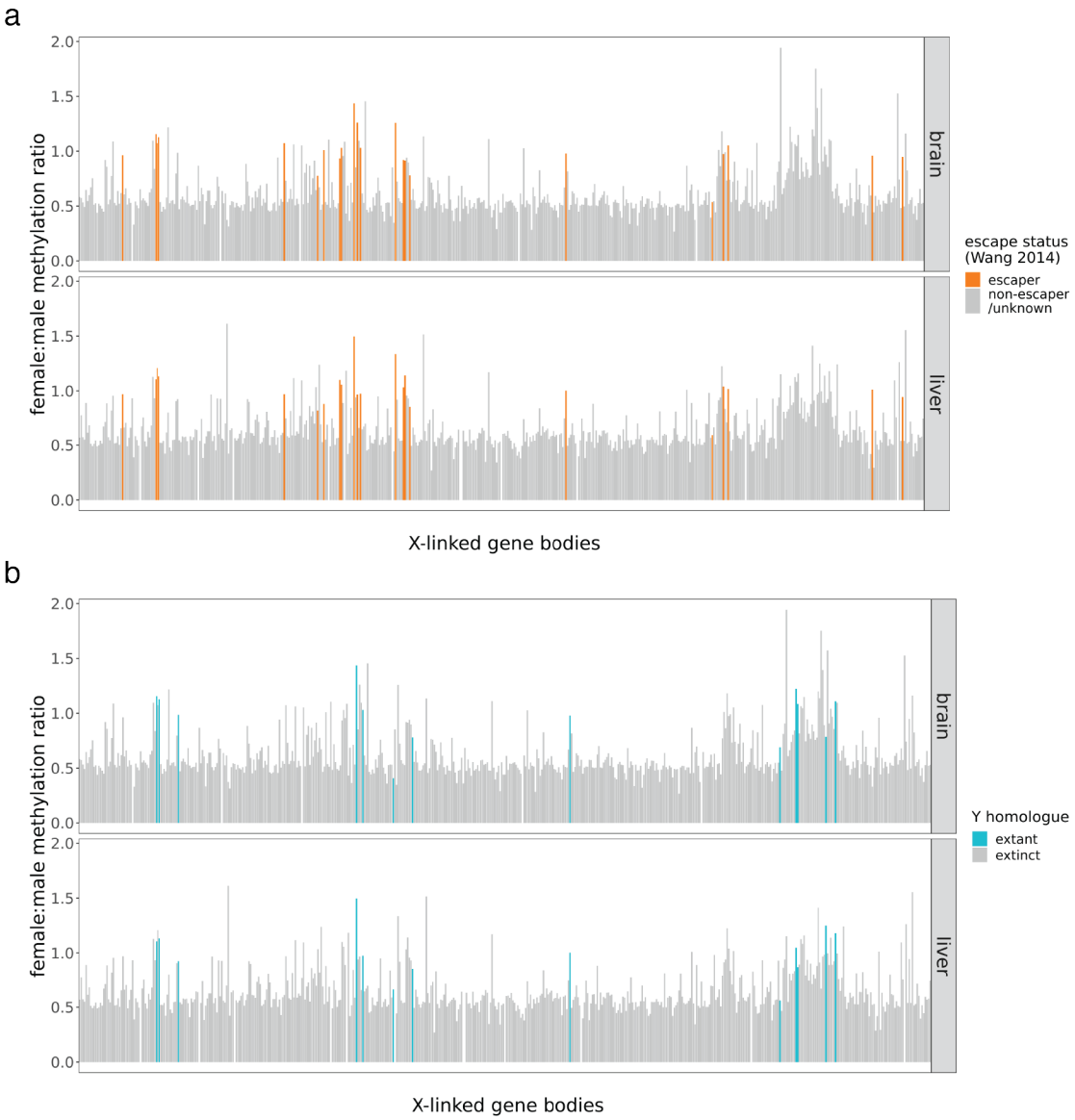


Figure 19. DNA methylation patterns of metatherian escape genes. a-b: Bar plots depicting the female:male methylation ratio of individual opossum X chromosome gene bodies (n = 533). Loci at which the methylation level was lower than 10% in one or both sexes were excluded from this analysis. **a**: Orange bars highlight genes previously reported to escape from X inactivation in opossum (Wang et al., 2014b). 22/24 reported escape genes are covered in this analysis. **b**: Blue bars highlight genes retaining an extant Y homologue in opossum (Cortez et al., 2014, Bellott et al., 2014). 13/18 genes retaining a Y homologue are covered in this data.

It has been suggested that X-linked genes retaining a Y homologue are likely to escape from XCI, as in this case, the effective dose in a male is 2, and there is no imperative for dosage compensation in the female (Jegalian and Page, 1998). This prediction would be expected to hold in the case of broadly-expressed Y genes that retain shared functionality with the X homologue, but not in the case of Y genes with restricted expression/that have

lost their ancestral function. In an RT-PCR analysis of Y gene expression 14/18 genes were broadly expressed (expressed in four or more tissues; Figure 20). Therefore, it might be expected that many opossum X-linked genes with an extant Y homologue would escape X inactivation.

Of the 18 X-linked genes retaining an extant Y homologue (Bellott et al., 2014, Cortez et al., 2014) 13 are covered in this data (blue bars, Figure 19b), and 6 of these genes were previously annotated as escape genes by Wang et al (2014b). The average female:male methylation ratio of those six genes previously annotated as XCI-escape genes is approximately 1 in both tissues (brain ratio: 1.08, liver ratio: 1.09; Table 4). For the seven genes with extant Y homologues that were not previously annotated as XC-escape genes, the female:male methylation ratio is approximately 0.9 (brain ratio: 0.9, liver ratio: 0.93), strongly suggesting they too escape from XCI (Table 4).

Three genes of the genes with extant Y homologues that were not previously annotated as escape genes, *HSFX*, *KLF8,* and *TFE3*, have a female:male methylation ratio lower than 0.8 in at least one tissue (*HSFX*: 0.41 in brain and 0.66 in liver; KLF8: 0.69 in brain and 0.56 in liver; *TFE3*: 0.76 in brain and 1.25 in liver; Table 4). These genes might therefore be tissue-specific escapers (especially in the case of *TFE3*, where the methylation levels in liver are strongly suggestive of an escape gene), or indeed not escape XCI. For *HSFX*, a gene where the female:male ratio is low in both brain and liver, the Y homologue displayed testis-restricted expression (*HSFY*; Figure 20). In addition, the Y-homologue of *KLF8*, *KLF8Y*, was expressed at a low level in several tissues, but was notably enriched in testis and spleen. These genes might therefore not be predicted to escape XCI, as the restricted expression of their Y-linked counterparts removes the requirement for dosage compensation.

In summary, these data show that metatherian escape genes display a unique methylation signature and lend support to the idea that X-linked genes with extant Y homologues are likely to escape X inactivation.

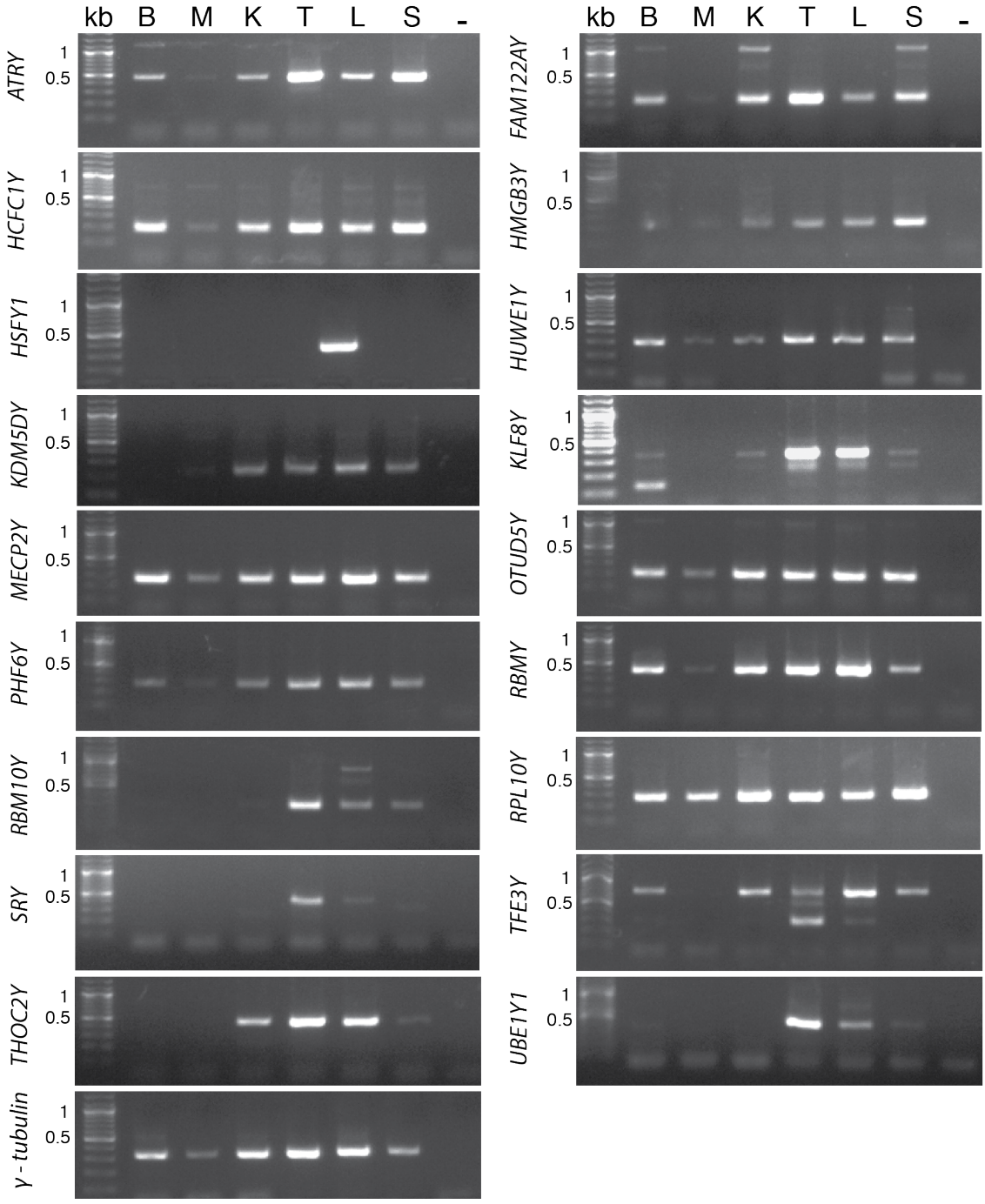


Figure 20. Expression patterns of metatherian Y-linked genes. Expression of 18 Y-linked genes by RT-PCR in adult male opossum tissues representing the three germ layers. *γ-tubulin* is included as a housekeeping control. B; brain, M; mandible, K; kidney, T: testis, L; liver, S: spleen, -; no template control. Individual bands were gel extracted and sequence validated by Sanger sequencing. Unexpected larger bands in *RBM10Y* and *FAM122AY* PCRs were Sanger sequenced and shown to constitute intronic sequence of the intended gene, indicating gDNA contamination or alternate splicing. Genes amplified in four or more tissues were considered broadly expressed. Shown are representative gels from one sample set; a biological duplicate produced the same result.

X gene	Y gene	Y gene expression pattern	Escape status (Wang *et al* 2014)	X gene covered in this analysis	Brain female:male methylation ratio	Liver female:male methylation ratio
ATRX	*ATRY*	broad	escaper	yes	0.98	1
FAM122A	*FAM122AY*	broad	escaper	yes	1.03	0.97
HCFC1	*HCFC1Y*	broad	escaper	yes	1.16	1.11
HMGB3	*HMGB3Y*	broad	escaper	yes	0.78	0.85
HSFX	*HSFY*	restricted		yes	0.41	0.66
HUWE1	*HUWE1Y*	broad		no		
KDM5c	*KDM5D*	broad		yes	1.11	1.18
KLF8	*KLF8Y*	broad		yes	0.69	0.58
MECP2	*MECP2Y*	broad	escaper	yes	1.13	1.13
OTUD5	*OTUD5Y*	broad		no		
PHF6	*PHF6Y*	broad		yes	1.44	1.5
RBM10	*RBM10Y*	restricted		yes	1.08	0.87
RBMX	*RBMY*	broad	escaper	no		
RPL10	*RPL10Y*	broad	escaper	no		
SOX3	*SRY*	restricted		no		
TFE3	*TFE3Y*	broad		yes	0.79	1.25
THOC2	*THOC2Y*	restricted		yes	0.99	0.93
UBE1	*UBE1Y1*	restricted		yes	1.22	1.04

Table 4: Methylation of metatherian XY homologues

3.7 Discussion

In this chapter, I used RRBS to profile the DNA methylation landscape of adult male and female mouse and opossum tissues. I adapted the RRBS technique to include digestion with a second enzyme, BfaI, chosen due its predicted increase in coverage of gene bodies and other genomic regions in an *in silico* RRBS model (Wang et al., 2013). However, it was necessary to sequence the libraries at a greater sequencing depth than typically utilised for MspI-only RRBS (Gu et al., 2011a) to attain coverage of the predicted number of CpG sites.

Global patterns of methylation in mouse and opossum RRBS libraries resembled the typical pattern reported for vertebrate species (Lister et al., 2009a), with low methylation at CGIs and high methylation across the remainder of the genome. Unlike genome-wide DNA methylation, X-chromosome methylation patterns were not conserved between mouse and opossum. The well-documented eutherian phenomenon of Xi CGI hypermethylation was evident in the mouse data presented here, an important validation of the method used. In contrast, there was no evidence for hypermethylation of CGIs on the opossum X chromosome. The findings presented here for metatherians replicate work from several previous reports (Kaslow and Migeon, 1987, Loebel and Johnston, 1996, Wang et al., 2014b, Waters et al., 2018) showing conclusively that female metatherians do not hypermethylate their X-linked CGIs.

Establishment of CGI hypermethylation on the eutherian X chromosome relies on the action of *Dnmt3b*, during the global remethylation of the embryonic genome (Auclair et al., 2014). A significant subset of CGIs are also reliant on the action of *Smchd1* for the correct establishment of methylation (Blewitt et al., 2008, Gendrel et al., 2012). It would be interesting to ascertain the expression pattern and functional activity of the opossum homologues of *Dnmt3b* and *Smchd1* during the establishment of opossum XCI; perhaps a difference in the expression timing of these genes accounts for the eutherian-metatherian differences in X chromosome CGI methylation.

I also confirmed and extended previous observations relating to methylation of non-CGI regions of the X chromosome. Waters et al. (2018) reported that the regions flanking hypomethylated promoters on the opossum X chromosome displayed differential methylation, concluding that the Xi was hypomethylated at these TSS-adjacent regions, and suggesting that this served a repressive function similar to the hypermethylated CGIs of eutherians. In examining the methylation of gene bodies and intergenic regions of the opossum X, I observed a pattern indicative of global hypomethylation of the Xi, rather than a specific TSS-related phenomenon. This finding is supported by earlier studies using immunofluorescence or nick-translation based approaches that found the Xi was globally hypomethylated in marsupials (Loebel and Johnston, 1993, Rens et al., 2010), and is concordant with various reports in eutherians of global hypomethylation of the Xi (Lock et al., 1986, Weber et al., 2005, Hellman and Chess, 2007, Gendrel et al., 2012, Cotton et al., 2015, Schultz et al., 2015, Keown et al., 2017, Duncan et al., 2018).

Metatherians therefore cannot rely on CGI hypermethylation for maintenance of XCI, implying that other epigenetic modifications are sufficient to maintain the silencing of this chromosome. For example, one study mapped the histone modifications H3K27me3 (‘repressive’) and H3K4me3 (‘active’) on the opossum X chromosome (Wang et al., 2014b). The active mark was enriched at X-linked promoters in males, with an intermediate level in females, while the repressive mark was depleted from transcribed genes on the male X chromosome. XCI-escape genes were found to have a profile similar to the active male X (Wang et al., 2014b).

The global differential methylation between Xa and Xi differs greatly in scale between opossum and mouse. In opossum, my data would suggest the Xi is essentially devoid of methylation, at the genomic features able to be profiled here, with the exception of the bodies of escape genes. In mice, by contrast, the difference between female and male X at intergenic and intragenic regions is approximately 10%.

These results raise interesting questions about the metatherian X chromosome. In eutherian mammals the genome is highly methylated, except during genome-wide reprogramming in the early embryo (Smith et al., 2012, Guo et al., 2013). It is interesting

to consider how DNA methylation may be excluded from the opossum Xi. It will be of great interest to investigate the methylation dynamics of the X chromosome in early opossum embryos.

DNA methylation is necessary for the successful suppression of repetitive elements housed in the genome. I have not explicitly examined repeat elements in this analysis, but it is interesting to consider that repeat elements may be hypomethylated to a similar degree as other regions of the opossum X chromosome. Repeat elements housed on the X chromosome might be at increased risk of activation, unless an alternative mechanism exists to ensure their suppression.

The presence of 5-methylcytosine increases the chance of a C to T transition mutation, which accounts for the general depletion of CpG sites in vertebrate genomes (Sved and Bird, 1990, Holliday and Grigg, 1993). The opossum X chromosome has a higher GC and CpG content than the autosomes (Mikkelsen et al., 2007b). It is possible that the globally hypomethylated Xi results in a lower overall mutation rate at X-linked cytosines, accounting for the higher GC content of the opossum X chromosome.

In both the opossum and the mouse, I observed the male X (and by assumption, the female Xa) to be slightly hypermethylated relative to autosomes. Given the suggested association between active transcription and gene body methylation, it is tempting to speculate that this may be related to X-upregulation (XCU), the process by which the male X and female active X are transcriptionally upregulated to achieve parity with the transcriptional output of the autosomes (Adler et al., 1997, Nguyen and Disteche, 2006, Gupta et al., 2006, Lin et al., 2011, Julien et al., 2012).

Several of the findings of this chapter could be unequivocally confirmed, and some of the questions raised could be addressed, by analysing the methylation of Xa and Xi separately, rather than as an average as done here. To address these points, in the next chapter I take advantage of mouse and opossum systems in which single-nucleotide polymorphisms (SNPs) can be used to analyse DNA methylation and gene expression in an allele-specific manner.

Chapter 4. Results 2

DNA methylation has important roles in regulation of gene expression, a notable example being the role of promoter hypermethylation in the maintenance of silencing on the Xi in female eutherian mammals (Mohandas et al., 1981, Graves, 1982, Csankovszki et al., 2001, Sado et al., 2000). In exception to this pattern, genes that in eutherian mammals escape XCI have promoters less methylated than those of their silenced neighbouring genes (Cotton et al., 2011, Sharp et al., 2011). Outside of the hypermethylated promoter/CGI regions, the eutherian Xi is slightly less methylated than its active counterpart (Weber et al., 2005, Gendrel et al., 2012, Keown et al., 2017, Duncan et al., 2018). Whether DNA methylation plays the same role in XCI in the metatherian lineage is not clear, as studies examining metatherian methylation have not observed promoter methylation on the Xi (Kaslow and Migeon, 1987, Loebel and Johnston, 1996, Wang et al., 2014b, Waters et al., 2018).

Initial studies of red blood cell antigens and activity levels of X-linked enzymes in women established that not every X-linked locus was silenced by XCI (Gorman et al., 1963, Fialkow, 1970, Shapiro, 1979). Cytological studies of replication timing suggested some regions of the human X chromosome might not be subject to XCI (Schempp and Meer, 1983). Subsequently, human/rodent hybrid cell lines carrying a single inactive copy of the human X chromosome were used to screen for genes that escaped XCI (Ellison et al., 1992, Yen et al., 1992, Wu et al., 1994, Carrel et al., 1999).

Comparison of human X-linked genes with their mouse orthologues showed that genes escaping XCI in humans were often subject to XCI in the mouse (Adler et al., 1991, Zinn et al., 1991, Disteche et al., 1992) though in other cases mouse orthologues were found to escape (Agulnik et al., 1994, Sheardown et al., 1996, Ehrmann et al., 1998). Next-generation sequencing technologies enabled comprehensive surveys of allele-specific gene expression across the eutherian X chromosome, facilitating the discovery of many novel escape genes, and revealed that approximately 3-7% of mouse, and 15% of human genes escape XCI (Yang et al., 2010, Berletch et al., 2015, Carrel and Willard, 2005, Zhang et al., 2013b, Cotton et al., 2013).

Initial studies of XCI in metatherians examined gene expression from a select few X-linked loci (Cooper et al., 1993, Samollow et al., 1989, Migeon et al., 1989). These studies found evidence of expression from the paternal allele, though the extent of expression varied between tissues and species profiled. It was therefore proposed that metatherian XCI was 'leaky', or prone to de-repression, a viewpoint supported by the absence of CGI methylation on the metatherian X chromosome (Cooper et al., 1993, Samollow et al., 1989, Migeon et al., 1989). However, subsequent studies using RNA fluorescence *in situ* hybridisation (RNA FISH) and RNA-seq found that metatherian XCI was efficient across a broad set of X-linked genes, contradicting the 'leaky' hypothesis (Mahadevaiah et al., 2009, Julien et al., 2012). A more recent examination of escape genes using allele-specific RNA-seq revealed that 24 genes escape XCI in the foetal brain and extra-embryonic membranes of the opossum (Wang et al., 2014b).

In Chapter 3, I profiled X chromosome DNA methylation in mouse and opossum tissues. In analysing mouse methylation, I replicated prior observations (Chapter 3 section 3.4). Opossum methylation profiles confirmed the absence of CGI methylation across the X chromosome, and in addition revealed an intriguing methylation pattern characterising opossum escape genes (Chapter 3 sections 3.5 and 3.6). I found that opossum genes known to escape XCI had a gene body methylation level approaching the male Xa, in contrast to other genes on the Xi, where the methylation level was depleted (Chapter 3 section 3.6). A significant number of genes not currently annotated as opossum escape genes also displayed this 'escape-methylation' signature, and I hypothesised these may represent novel escape genes.

However, these analyses relied on interpreting the average of the Xa and Xi in female samples, resting on the assumption that the female Xa would resemble the single male Xa. I therefore aimed to extend this work by analysing methylation on the Xa and the Xi separately. Given the observations regarding escape genes made in the previous chapter, I also aimed to characterise gene expression from the Xa and Xi and integrate this data to investigate the relationship between methylation and gene expression on the X chromosome.

Analysis of gene expression or epigenetic profiles of the Xa and Xi individually can be achieved by sequencing samples in which the two alleles can be differentiated by the presence of single nucleotide polymorphisms (SNPs). Such an approach has been employed previously in mouse using cultured cells (Yang et al., 2010) and brain and spleen tissue (Berletch et al., 2015). In opossum, a similar survey of XCI-escape genes has been performed in foetal brain and extra-embryonic membranes, but not in adult tissues (Wang et al., 2014b).

In this chapter I analyse the methylation profile of mouse and opossum X chromosomes using allele-specific BS-seq, allowing the Xa and the Xi to be examined separately. Complementing the BS-seq dataset with an allele-specific analysis of gene expression, I identify XCI-escape genes, in order to examine the relationship between DNA methylation and escape from XCI.

4.1 Generation of allele-specific datasets

4.1.1 Mouse and opossum crosses used in this analysis

Allele-specific analysis of RNA-seq and BS-seq required that aligned sequence reads be assigned to their parental allele of origin. In order to achieve this, genetically diverse parents were crossed and RNA-seq and BS-seq libraries were generated from somatic tissues of the male and female offspring of these crosses. Variants between the parental genomes facilitated allele-specific analysis of the resultant datasets.

In this analysis, the publicly available software tool SNPsplit was used to perform read assignments to parental alleles based on the presence of variants between the parental genomes (Krueger and Andrews, 2016). The SNPsplit pipeline utilises SNPs that differ between parental genomes but is not equipped to tolerate heterozygous positions within a parental genome. Therefore, only those loci that were homozygous (or hemizygous in the case of the male X chromosome) for different alleles within each parent could be utilised in this analysis.

Random XCI in the mouse soma precludes allele-specific analysis of the X chromosome. I therefore utilised a C57BL/6J *Xist*^{tm1Jae} x *Mus spretus* F1 hybrid cross. The *Xist*^{tm1Jae} (*Xist*$^{-}$) allele is an approximately 15 kb deletion that abolishes *Xist* function and results in complete skewing of XCI (Marahrens et al., 1997). In heterozygous female offspring from this cross, the C57BL/6J X chromosome is active, and the *M. spretus* X chromosome is inactive (Figure 21a). I generated BS-seq and RNA-seq libraries from brain and liver tissue of three heterozygous female offspring, and from three of their *Xist*$^{-}$ male littermates. Both parents of the C57BL/6J x *M. spretus* cross are inbred strains for which genetic variation is well characterised and publicly available. 41,668,158 SNPs between the C57BL/6J and *M. spretus* genomes were accessed from the Mouse Genomes Project genetic variation dataset (Keane et al., 2011; Figure 22a).

Opossum XCI is imprinted, always affecting the paternal X chromosome, and is therefore amenable to allele-specific analysis without requiring artificial skewing of XCI (Figure 21b). However, it was not possible to generate an interspecific opossum cross because inbred opossum strains were not available. To maximise the available genetic variation, a cross was set up using parent animals from the LL2 opossum stock (collaboration with John VandeBerg, University of Texas; Figure 21b). The LL2 stock was derived from founder animals trapped in geographically varied regions of Brazil and is expected to harbour greater genetic variation than the Francis Crick Institute opossum colony (John VandeBerg, personal communication). I generated RNA-seq and BS-seq libraries from the brain, liver, and spleen of three female and three male offspring of this cross.

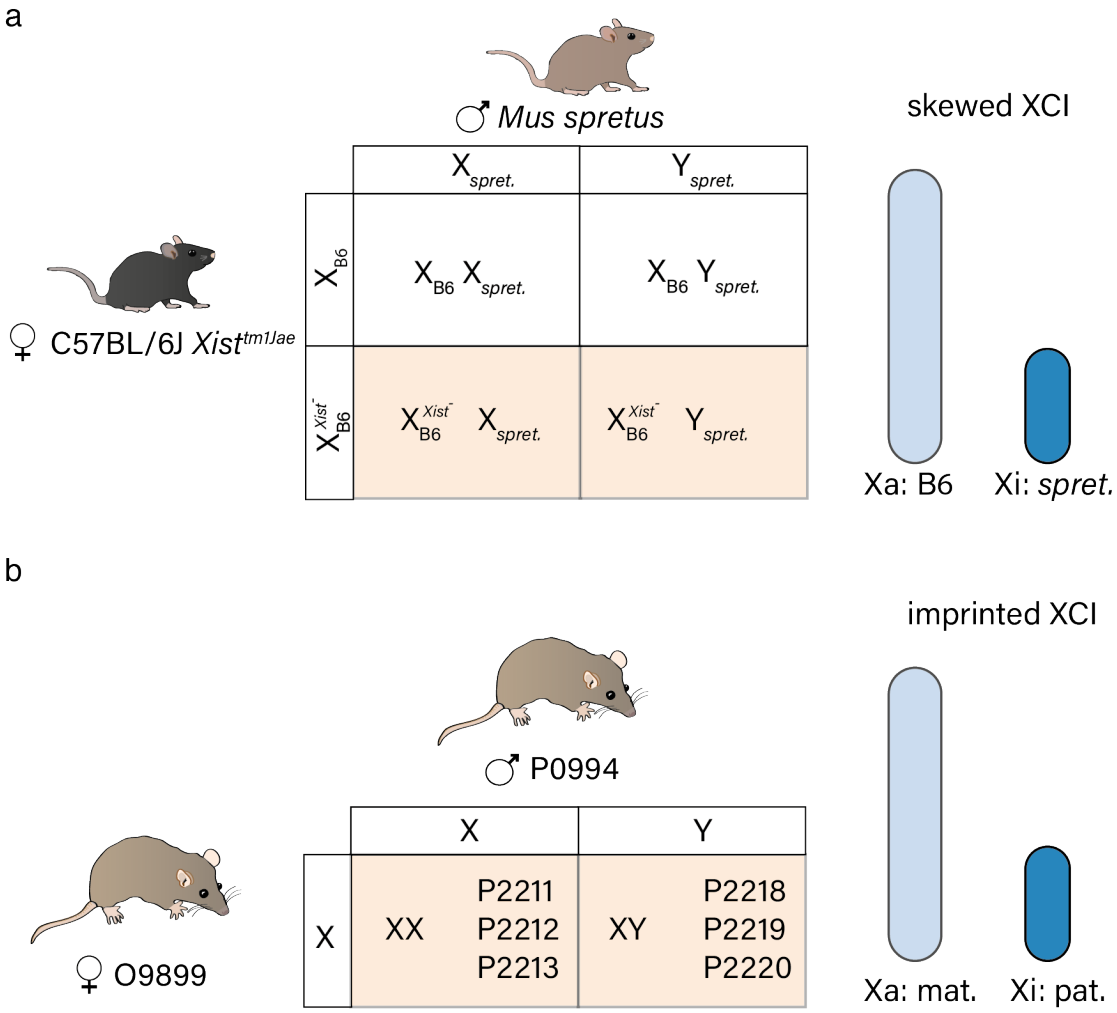


Figure 21: Description of mouse and opossum samples used in this chapter. Diagrams depicting the mouse and opossum crosses used for allele-specific RNA-seq and BS-seq analyses in this chapter. **a**: Mouse: C57BL/6J *Xist*^{tm1Jae} x *Mus spretus* hybrid cross. Orange shaded boxes indicate the F1 genotypes from which libraries were generated. Female offspring carrying the *Xist*^{tm1Jae} (*Xist*$^{-}$) allele have skewed XCI such that the maternally inherited B6 X chromosome is active and the paternally inherited *M. spretus* X chromosome is silenced. **b**: Opossum: cross of individuals from the LL2 stock maintained by John VandeBerg, University of Texas. Orange shaded boxes indicate the specific offspring from which libraries were generated. In female opossums, imprinted XCI results in an active maternal and inactive paternal X chromosome.

Genetic variation in the LL2 opossum population has never been analysed in a genome-wide manner (John VandeBerg, personal communication). In order to define a list of SNPs that differed between the opossum parental genomes, whole-genome sequencing (WGS) of female O9899 and male P0994, followed by a *de novo* discovery of genetic variants in the two genomes was performed by Jasmin Zohren (Turner laboratory).

25,169,691 variants were discovered, of which 20,966,534 were SNPs. 2,077,304 of the SNP positions were homozygous in each parent and thus useable in the SNPsplit analysis (Figure 22a).

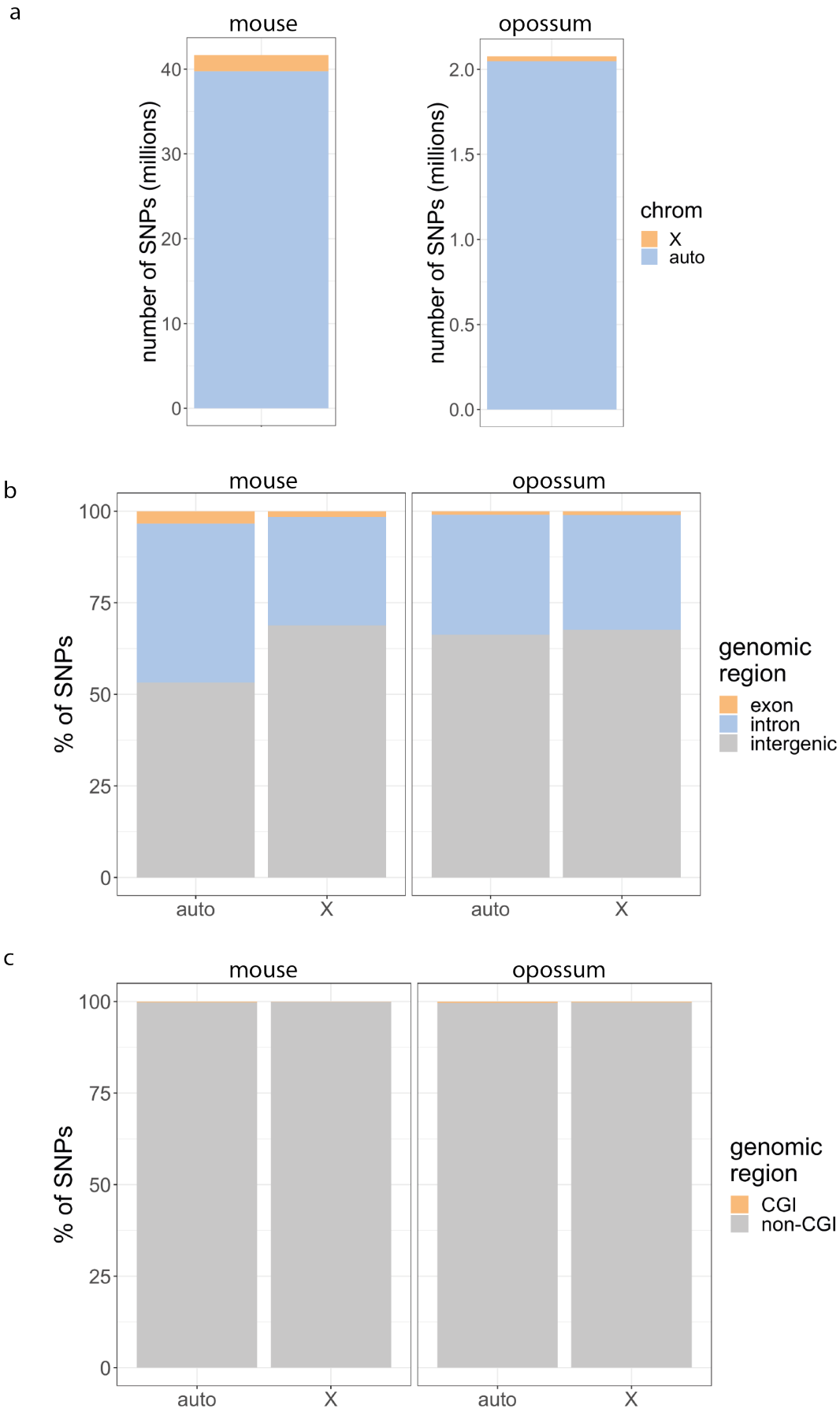

a
mouse
opossum
number of SNPs (millions)
40
30
20
10
0
2.0
1.5
1.0
0.5
0.0
chrom
X
auto
b
mouse
opossum
% of SNPs
100
75
50
25
0
auto
X
genomic region
exon
intron
intergenic
c
mouse
opossum
% of SNPs
genomic region
CGI
non-CGI

Figure 22: Genomic distribution of SNPs in mouse and opossum crosses. a: Number of SNPs on autosomes and X chromosomes of C57BL/6J x *Mus spretus* (mouse), or LL2 (opossum) samples. **b-c**: Stacked bar charts showing the proportion of SNPs found in different genomic regions of C57BL/6J x *Mus spretus* (mouse), or LL2 (opossum) samples. **b**: SNP distribution in exons, introns, and intergenic regions. Exon: Ensembl defined exons, excluding those from pseudogenes; intron: Ensembl defined introns, excluding those from pseudogenes; intergenic: any region of the genome not encompassed by the exon and intron categories. **c**: SNP distribution in UCSC-defined CpG islands (CGI), or regions outside CGIs. *De novo* discovery of genetic variants in the opossum was performed by Jasmin Zohren, Turner laboratory.

There were far fewer SNPs in the opossum than in the mouse (Figure 22a). I calculated the percentage of SNPs found in different genomic features and found a slight difference in the genomic distribution of SNPs in the two species (Figure 22b-c). A lower proportion of SNPs were found in exonic and intronic regions in opossum than in mouse. There are several potential explanations for this difference. Firstly, functional coding sequences are expected to be constrained, and this effect would be stronger in closely related animals from the same population, like the opossum parents, resulting in fewer variants within genes. Secondly, while there are similar numbers of protein-coding genes annotated in both animals (mouse = 21,900; opossum = 21,328, Ensembl defined genes), there are nearly twice as many lncRNA genes annotated in the mouse (mouse = 10209, opossum = 6472). Therefore, it is possible that some opossum SNPs are contained within as yet unannotated genes. Thirdly, the opossum genome is larger (mouse = ~2.7 Gb, opossum = ~3.6 Gb, Ensembl golden path length), but with approximately the same number of genes, meaning the intergenic space in the opossum genome is greater, and can harbour relatively more SNPs.

The differences in SNP number and distribution implied that the opossum allele-specific analysis would not achieve the same genomic scope as the mouse analysis (mouse: ~ 1 SNP/70 bp; opossum: ~ 1 SNP/1.8 Kb on average over the genome). However, with 2,077,304 potentially informative SNPs, I concluded that the genomic coverage achieved in the opossum would permit an examination of chromosome-wide trends relating to XCI, and therefore proceeded to analyse XCI escape and DNA methylation patterns in these datasets.

4.1.2 **Generation of RNA-seq datasets**

RNA was purified from the brain, liver, and spleen of opossum samples, representing different developmental germ layers. Samples from mouse brain and liver were also processed to act as positive control samples. RNA quality was assessed by TapeStation analysis (Figure 23a-b). Samples with an RNA Integrity Number (RIN) between 6.8 – 10 were processed by the Francis Crick Institute ASF to produce RNA-seq libraries (Figure 23c; Table 5).

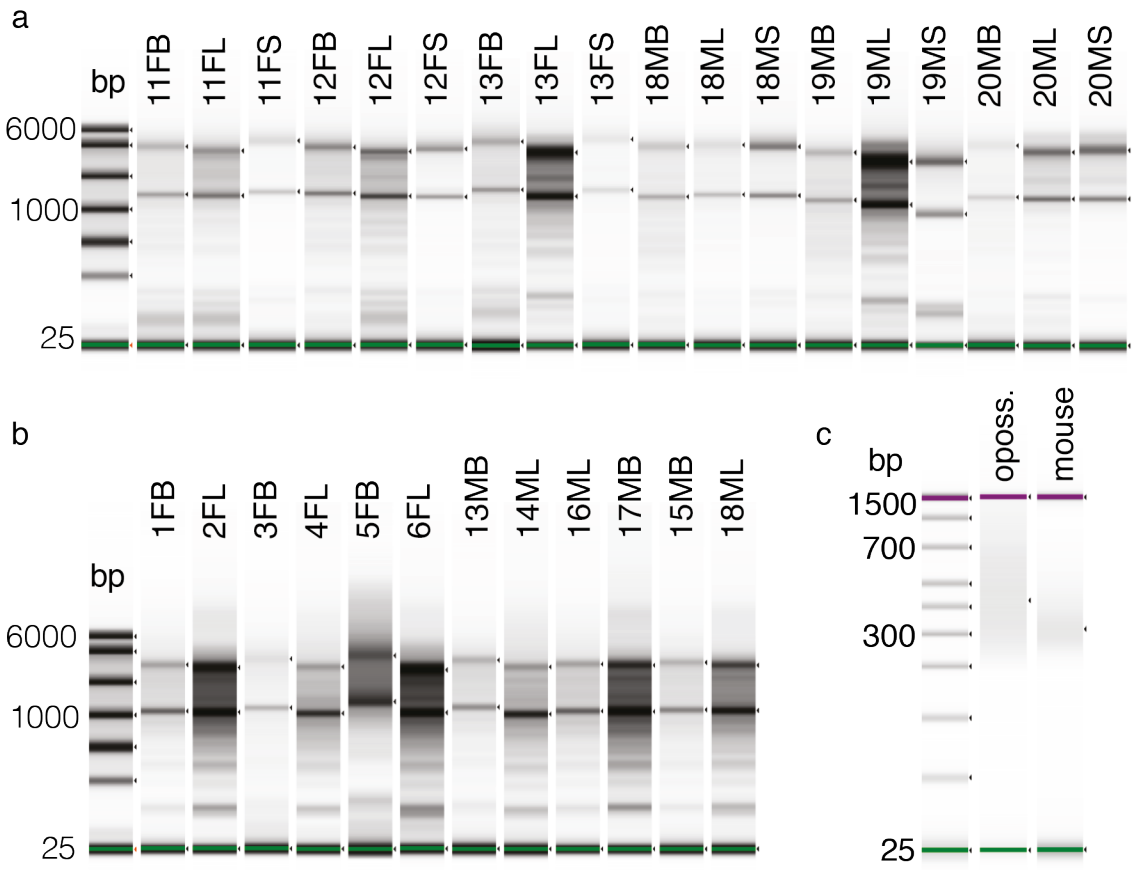


Figure 23: Generation of RNA-seq libraries. RNA-seq libraries were prepared from the brain, liver, and spleen of three male and three female opossums, and from the brain and liver of three male and three female mice. **a-b**: TapeStation e-gel profiles of purified RNA. **a**: opossum RNA, **b**: mouse RNA**. c**: Example TapeStation e-gel profiles of opossum and mouse RNA-seq libraries. RNA-seq libraries were prepared from purified RNA by the Francis Crick Institute Advanced Sequencing Facility.

The resulting libraries were sequenced paired-end, attaining read depths of between 54 - 156 million (Table 5). Approximately 90% of bases were retained in mouse libraries after adapter and quality trimming, while in opossum libraries, approximately 95% of bases were retained (Table 5). All libraries had phred scores above 30 (Figure 24a). After alignment to the reference genome with HISAT2, opossum libraries had mapping rates between 65-71%. Mouse libraries had mapping rates between 45-56%.

In the mouse RNA-seq libraries, 50-65 % of reads encompassed a SNP and could be assigned to parental alleles by SNPsplit. In the opossum libraries ~ 2% of reads were assignable to a parental allele (Figure 24b). As a quality control measure, the ratio of reads assigned to parental genomes was calculated for all genes at which any allelic reads had been assigned. Approximately 50% of reads were assigned to each parent in all libraries, as would be expected given that the majority of gene expression is anticipated to be biallelic (Figure 24c). There was wider variation in the paternal/maternal ratio of read assignment in opossum than in mouse, reflective of the lower total number of reads producing 'noisy' ratios. Following this observation, I decided to apply a minimum threshold of allele-assigned reads in the subsequent analysis of allelic ratios at specific gene loci (Chapter 4 section 4.2).

Sample	Species	RIN	Number of reads	Bases retained after filtering (%)	Mapping rate (%)
1M Brain	Mouse	8.3	156,515,538	90.40%	49.26%
2M Liver	Mouse	7.6	87,943,746	89.05%	45.89%
3M Brain	Mouse	7.8	87,249,922	91.40%	51.81%
4M Liver	Mouse	7.4	81,360,478	90.20%	49.10%
5M Brain	Mouse	7.6	88,454,236	90.10%	48.14%
6M Liver	Mouse	7.6	95,266,358	90.45%	51.09%
13F Brain	Mouse	7.7	78,721,584	90.20%	47.97%
14F Liver	Mouse	7.5	82,186,242	91.15%	51.10%
15F Brain	Mouse	7.5	91,652,538	89.50%	46.10%
16F Liver	Mouse	7.3	91,217,472	91.40%	52.91%
17F Brain	Mouse	7.7	84,399,662	93.10%	56.73%
18F Liver	Mouse	7.5	83,699,956	91.25%	51.55%
11F Brain	Opossum	7.7	64,652,218	94.52%	65.29%
11F Liver	Opossum	8.3	76,111,590	95.78%	67.43%
11F Spleen	Opossum	9.6	73,784,504	93.98%	65.74%
12F Brain	Opossum	8.5	58,886,864	96.63%	71.46%
12F Liver	Opossum	8.4	93,212,104	96.62%	68.33%
12F Spleen	Opossum	10	55,837,402	95.37%	71.05%
13F Brain	Opossum	8	145,421,204	96.80%	71.29%
13F Liver	Opossum	8.6	54,803,334	96.43%	70.50%
13F Spleen	Opossum	8.1	75,134,594	95.27%	68.40%
18M Brain	Opossum	7	61,560,880	96.45%	69.71%
18M Liver	Opossum	8.4	75,101,216	96.65%	69.83%
18M Spleen	Opossum	9.8	57,709,466	96.37%	72.50%
19M Brain	Opossum	6.9	63,445,082	96.40%	69.47%
19M Liver	Opossum	8.8	37,984,144	95.77%	70.54%
19M Spleen	Opossum	9.5	54,109,292	96.13%	71.51%
20M Brain	Opossum	6.8	68,307,854	96.53%	69.75%
20M Liver	Opossum	9.3	63,682,662	96.40%	71.42%
20M Spleen	Opossum	9.9	60,368,628	96.15%	71.72%

Table 5: RNA-seq library statistics

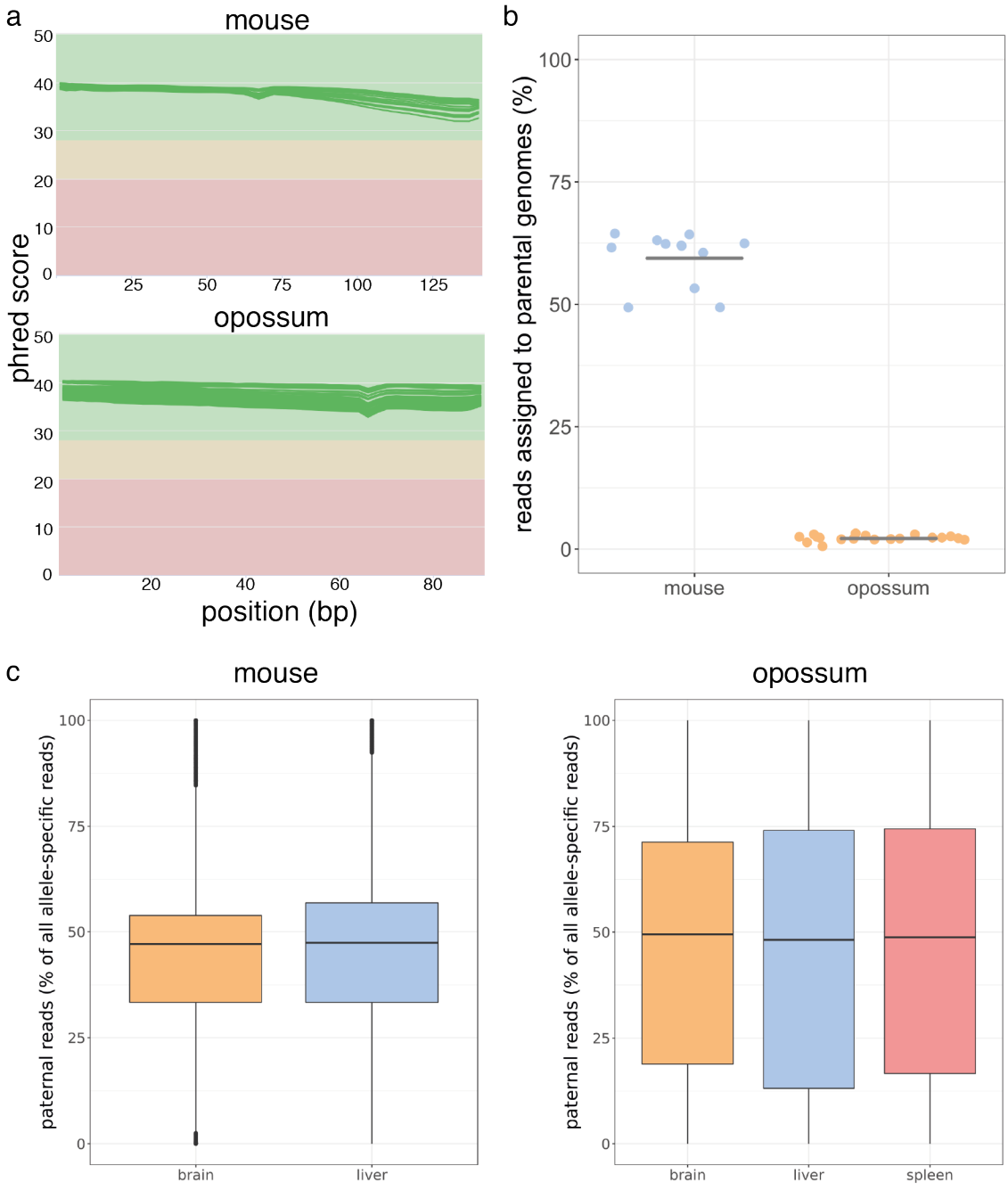


Figure 24: Quality control of allele-specific RNA-seq datasets. a: Per base sequence quality plots. **b**: Percentage of the total mapped reads per library assignable to a parental genome. **c**: Boxplots showing paternal reads as a percentage of the total allele-specific reads per gene, for all genes for which allele-specific reads were detected.

4.1.3 **Generation of BS-seq datasets**

High molecular weight genomic DNA was purified from the same mouse and opossum tissue samples used to generate the RNA-seq libraries described above (Figure 25a-b). These DNA samples were used to construct BS-seq libraries following the method described for bulk samples in Clark et al. (2017). Quality control TapeStation analyses determined that the resulting libraries had size distributions within the expected range (Figure 25c-d).

The BS-seq libraries were sequenced paired end, attaining read depths of between 198-363 million (Table 6). Approximately 90% of base pairs were retained following adapter and quality trimming (Table 6). All libraries had phred scores above 30 (Figure 26a), and mapping rates between 53-67% (Table 6). The average CHH methylation rate in these libraries was 1.64%, implying a bisulfite non-conversion rate of less than 2% (Table 6).

The average methylation rate of cytosines in a CpG context was 64.5%, though opossum spleen samples had slightly lower CpG methylation ratios than the other tissues, ranging from 49-60%. In the mouse BS-seq libraries, between ~ 65 % of reads were able to be assigned to parental alleles by SNPsplit. In opossum libraries ~ 2% of reads were assignable to a parental allele (Figure 26b).

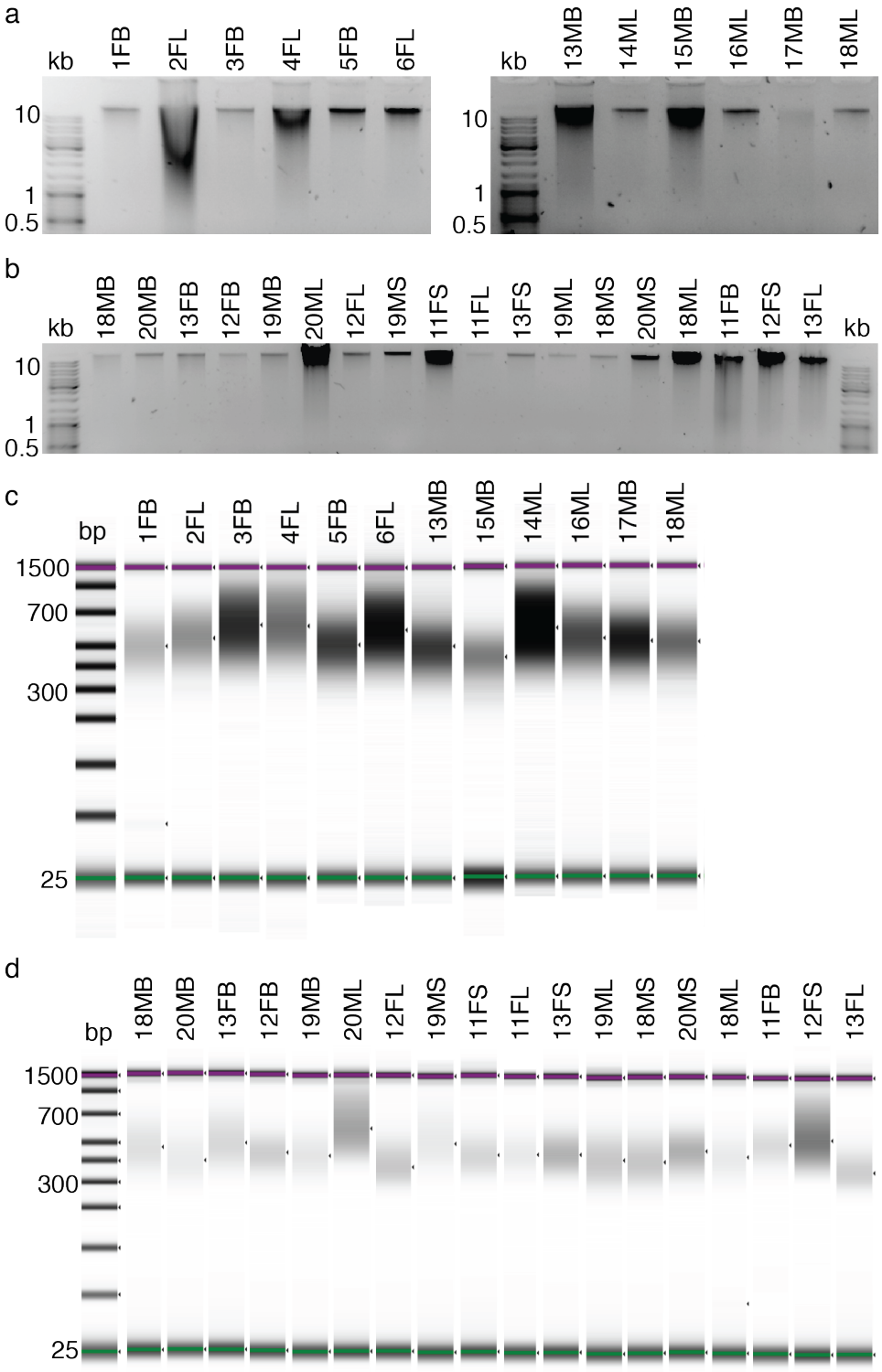


Figure 25: Generation of BS-seq libraries. BS-seq libraries were prepared from the brain, liver, and spleen of three male and three female opossums, and from the brain and liver of three male and three female mice. **a-b**: High molecular weight genomic DNA. **a**: Mouse gDNA, **b**: Opossum gDNA. **c-d**: TapeStation e-gel profiles of resultant BS-seq libraries. **c**: Mouse libraries, **d**: Opossum libraries. TapeStation analyses were performed by the Francis Crick Institute Advanced Sequencing Facility.

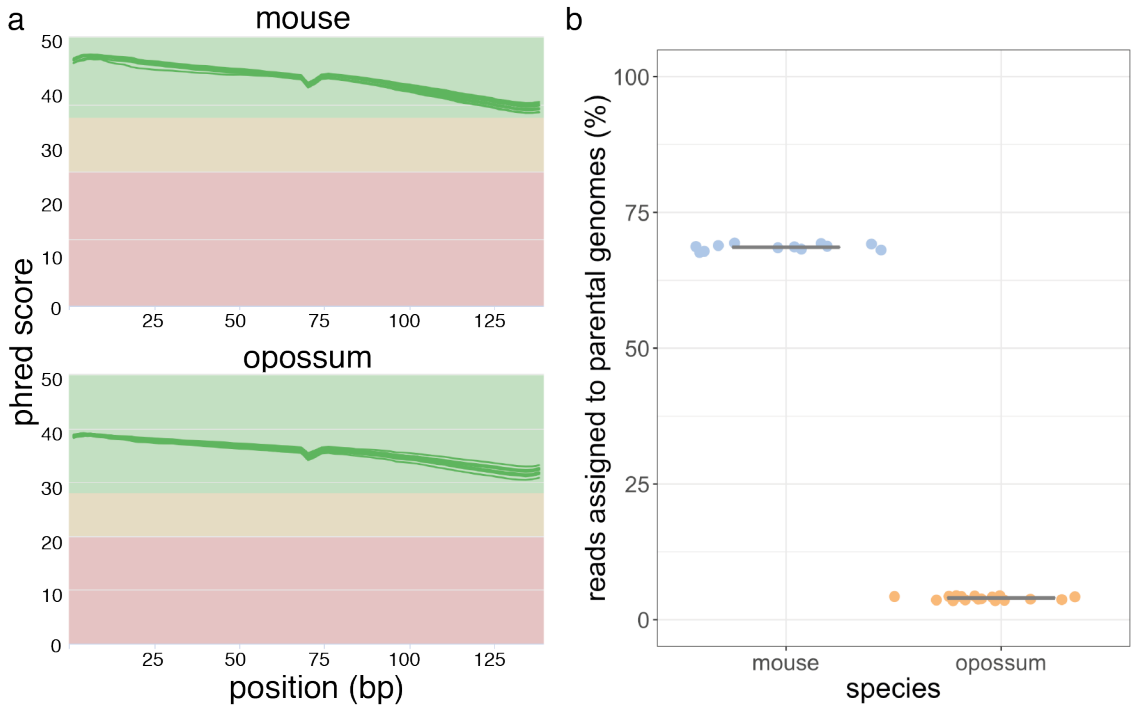


Figure 26: Quality control of allele-specific BS-seq datasets. a: Per base sequence quality plots. **b**: Percentage of the total mapped reads per library assignable to a parental genome; grey bar indicates the mean.

Following these QC checks, final datasets were prepared for analysis by filtering for coverage. In the mouse, coverage filtering was applied according to the following criteria: CpG sites covered by a minimum of three reads in at least two biological replicates in all tissues were retained for analysis. Subsequently, biological replicates were pooled, resulting in a minimum read coverage of six reads per site. These filtering conditions resulted in a final set of 1,480,274 CpGs represented in both tissues.

In the opossum, because fewer reads were able to be allele-assigned, I chose to implement a more lenient coverage filter, in order to retain more CpG sites for analysis. Sites covered by a minimum of two reads in at least two biological replicates were retained for analysis, resulting in a minimum read coverage of four reads per site upon pooling of biological replicates. These filtering conditions resulted in a final set of 104,992 CpGs in the brain, 31,763 CpGs in the liver, and 125,740 CpGs in the spleen.

Having successfully generated allele-specific RNA-seq and DNA methylation datasets, I proceeded to characterise gene expression from the active and inactive X in mouse and

opossum (section 4.2). This expression data was then integrated into my subsequent analysis of X-chromosome DNA methylation (sections 4.3 and 4.4).

Sample	Species	Number of reads	Bases retained after trimming (%)	Mapping rate (%)	C methylated in CHH context (%)	C methylated in CpG context (%)
1M Brain	Mouse	206,698,936	91.50%	57.40%	2.80%	68.90%
2M Liver	Mouse	274,160,268	92.10%	56.60%	1.40%	70.80%
3M Brain	Mouse	222,637,534	92.00%	53.20%	2.50%	70.80%
4M Liver	Mouse	255,642,248	92.10%	54.80%	1.30%	70.30%
5M Brain	Mouse	212,509,872	90.20%	55.40%	3.00%	69.10%
6M Liver	Mouse	363,935,682	91.90%	54.50%	1.50%	69.20%
11F Brain	Opossum	201,993,022	90.10%	57.60%	1.90%	65.50%
11F Liver	Opossum	200,762,004	95.90%	63.20%	1.10%	66.50%
11F Spleen	Opossum	215,094,862	95.20%	67.20%	1.00%	60.50%
12F Brain	Opossum	205,623,286	95.40%	60.80%	2.00%	68.10%
12F Liver	Opossum	209,631,620	94.40%	56.50%	1.40%	61.50%
12F Spleen	Opossum	331,246,282	85.10%	58.00%	1.10%	49.20%
13F Brain	Opossum	223,265,050	94.30%	61.60%	1.50%	68.10%
13F Liver	Opossum	198,846,600	93.50%	54.30%	1.90%	61.00%
13F Spleen	Opossum	206,531,870	95.70%	60.70%	1.20%	55.40%
18M Brain	Opossum	200,254,520	93.80%	61.90%	1.50%	68.40%
18M Liver	Opossum	202,729,910	91.70%	57.80%	1.30%	64.00%
18M Spleen	Opossum	202,113,488	91.70%	59.00%	1.20%	54.60%
19M Brain	Opossum	208,937,604	93.70%	60.90%	1.70%	68.70%
19M Liver	Opossum	206,133,846	92.20%	58.70%	1.10%	61.50%
19M Spleen	Opossum	210,188,946	94.10%	62.20%	0.90%	54.00%
20M Brain	Opossum	205,434,532	93.30%	63.10%	1.60%	69.70%
20M Liver	Opossum	268,034,356	93.20%	61.00%	1.10%	69.50%
20M Spleen	Opossum	209,307,760	89.90%	56.90%	1.30%	54.70%

Table 6: BS-seq library statistics

4.2 Detection of XCI-escape genes using allele-specific analysis of RNA-seq

In order to interpret the methylation profiles of opossum and mouse Xa and Xi, I sought first to characterise X-linked genes as subject to or escaping from XCI, using allele-specific RNA-seq. I first aimed to reproduce previous findings regarding mouse escape genes in the brain (Berletch et al., 2015). I also extended this analysis to mouse liver, as this was a tissue in which I had profiled DNA methylation, and for which there was no previous survey of escape genes. Following validation of my approach using mouse data, I aimed to characterise escape genes in opossum adult tissues.

To be categorised as an escape gene in this analysis, I considered that a gene must a) be expressed in the relevant tissue with an FPKM >1, b) encompass at least one SNP and have a minimum of 10 allele-assigned reads to avoid lowly-covered genes, which could suffer from 'noisy' ratios, and c) have more than 10% expression from the paternal (Xi) allele, i.e. an allelic ratio of $reads_{paternal}$ / ($reads_{paternal}$ + $reads_{maternal}$) > 0.1. As a quality control, I excluded any genes with an allelic ratio > 0.1 in male samples, as these represent false positive genes likely to derive from mismapping of autosomal sequences.

4.2.1 Mouse XCI-escape genes

Using the complete set of mapped RNA-seq reads, FPKM values were calculated for each tissue. 13,375 genes were expressed at FPKM >1 in brain, while 11,769 genes were expressed in liver. 10,323 of these genes were expressed in common between the tissues. Of these expressed genes, 13,325 had reads assigned to parental alleles in the brain, and 11,704 in the liver. Application of a minimum threshold of ten allele-specific reads in female resulted in 13,223 brain-expressed genes and 11,491 liver-expressed genes being retained for analysis (99% and 98% of expressed genes, in brain and liver respectively; Figure 27a).

Autosomal and X chromosome allelic expression ratios were calculated for each gene (Figure 27b). Most genes on autosomes were expressed at approximately equal levels

from each allele (allelic ratio = 0.477, Figure 27b). In contrast, most X-chromosome genes were expressed exclusively from the Xa (allelic ratio = 0.00514, Figure 27b). Nine genes in the brain and 11 genes in the liver did not express exclusively from the Xa (allelic ratio > 0.1) and were thus identified as escaping XCI (Figure 27c; Appendix 7.4).

A previous survey of mouse escape genes in brain, spleen, and ovary categorised 12 genes as escapers common to two or more tissues (Berletch et al., 2015). In the current analysis, I identified eight escape genes in common between brain and liver (Figure 27d). All eight genes had previously been reported to escape XCI in multiple tissues (Berletch et al., 2015, Tian et al., 2010).

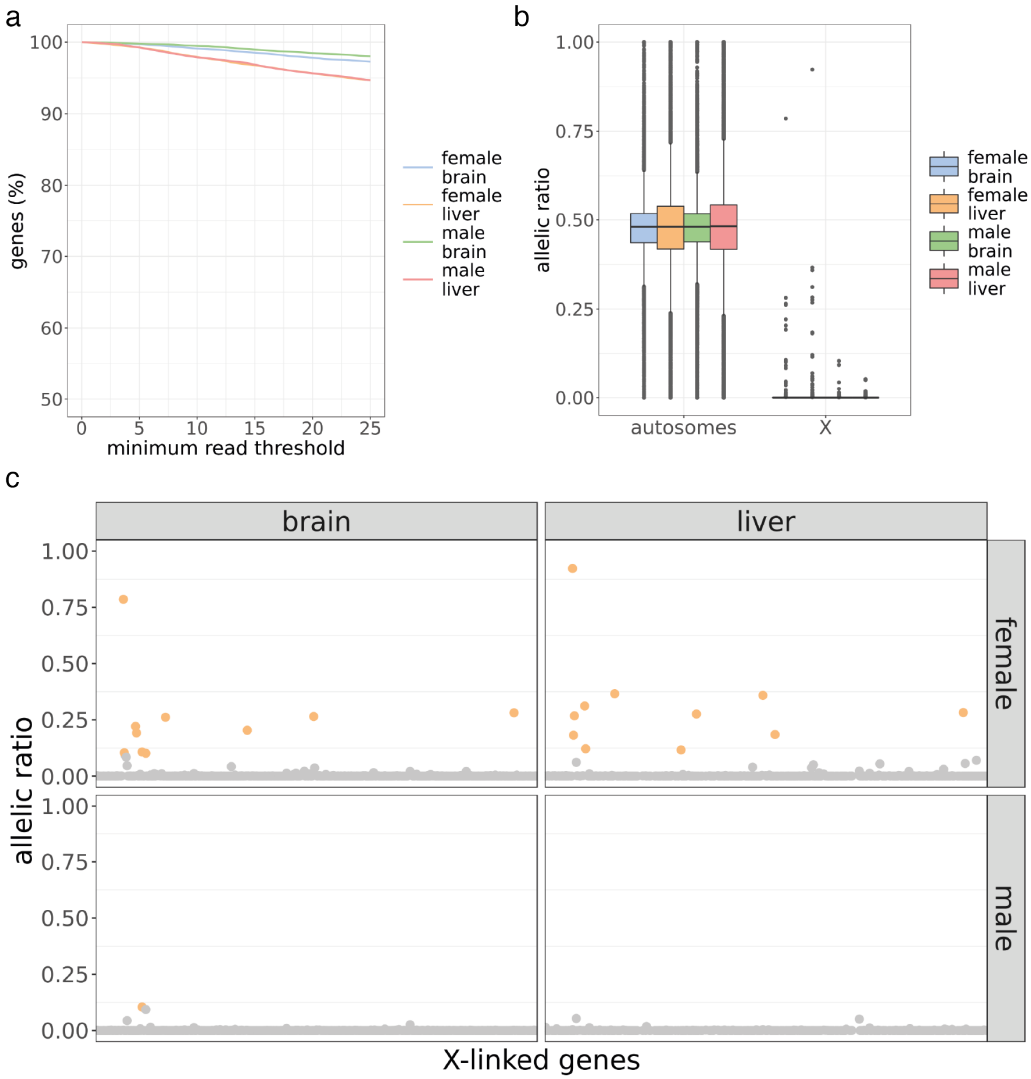

a
genes (%)
minimum read threshold
female brain
female liver
male brain
male liver
b
allelic ratio
autosomes
X
c
brain
liver
female
male
allelic ratio
X-linked genes
d
brain (9)
liver (11)
Lpar4†
Ddx3x*
Kdm5c*
Pdbdc1*
Eif2s3x*
Kdm6a*
Xist*
5530601H04Rik*
Jpx‡
Tmem29*
Ftx*
Cfp*

Figure 27: Allele-specific gene expression in mouse. a: Line plots showing the percentage of genes with FPKM > 1 retained for analysis under different minimum read thresholds. **b-c**: Allelic expression ratios of genes with FPKM > 1 and >9 allele-specific reads in females. Allelic expression ratios were calculated as $reads_{paternal}/(reads_{paternal} + reads_{maternal})$. $n_{brain} = 13223$; $n_{liver} = 11491$. **b**: Boxplots of allelic expression ratios for the autosomes and X chromosome. **c**: Allelic expression ratios of X-linked genes. Orange-coloured points indicate genes with an allelic expression ratio >0.1. $n_{brain} = 454$; $n_{liver} = 325$. **d**: Venn diagram depicting the XCI escape genes detected in brain and liver. Citations for genes previously reported to escape XCI are denoted by superscripts: *Berletch et al. (2015), †Wang et al. (2017), ‡Tian et al. (2010). Genes in bold were previously described as spleen-specific XCI escape genes (Berletch et al., 2015).

In addition to 'common' escape genes, the previous survey had categorised 26 genes as tissue-specific XCI escapers (Berletch et al., 2015). In the current analysis I identified four tissue-specific escape genes; one in brain, and three in liver. Of the four genes, *Ftx* and *Lpar4* had previously been reported as XCI escapers in multiple other tissues (Berletch et al., 2015, Wang et al., 2017). The two other genes, *Cfp* and *Tmem29*, had previously been categorised as tissue-specific escapers (Berletch et al., 2015). *Cfp* and *Tmem29* can now therefore be considered 'common' escapers, as I have shown that they escape XCI in a second tissue.

Having replicated most of the previously identified 'common' mouse escape genes, I considered that my methodology was successful, and proceeded to apply the same approach to opossum allele-specific RNA-seq.

4.2.2 **Opossum XCI-escape genes**

I calculated FPKM values for each tissue using the complete set of mapped RNA-seq reads. 13,986 genes were expressed at FPKM >1 in brain, 12,070 in liver, and 12,265 in spleen. 10,379 of these genes were expressed in common between the tissues. Of these expressed genes, 5,769 had reads assigned to parental alleles in the brain, 4,046 in the liver, and 4,167 in spleen. Application of a minimum threshold of ten allele-specific reads in female resulted in 3,537 brain-expressed genes, 2,232 liver-expressed genes, and 2,497 spleen-expressed genes being retained for analysis (61%, 55%, and 59% of expressed genes, in brain, liver and spleen respectively; Figure 28a).

Examination of allelic ratios showed that autosomal genes were expressed biallelically (allelic ratio = 0. 482; Figure 28b), while X-chromosome genes were mostly expressed exclusively from the Xa (allelic ratio = 0.0250, Figure 28b). Three genes in the brain, one gene in the liver, and six genes in the spleen were also expressed from the inactive X and classified as escape genes (allelic ratio > 0.1; Figure 28c).

A previous survey identified 24 common escape genes in opossum foetal brain and extra-embryonic membranes (Wang et al., 2014b). My analysis identified one escape gene, *RBMX,* in common between all three tissues profiled (Figure 28d). Of the escape genes I detected here, only *RBMX* and the spleen escape gene *FLNA* were among the 24 escape genes previously identified (Wang et al., 2014b).

The previous survey did not identify any tissue-specific escape genes (Wang et al., 2014b). Therefore in my analysis the genes *ZBTB33, G-protein coupled receptor 12-like*, *ARHGAP4, GLA, RRAGB,* and ENSMODG00000025578 (a protein coding gene with homology to *Drosophila melanogaster RpL36A*), were identified as novel tissue-specific opossum escape genes (Figure 28d; Appendix 7.4).

I successfully replicated two known escape genes (*RBMX*, *FLNA*) and in addition identified several novel escapers. It was therefore possible to characterise eight X-linked genes as escapers based on my analysis. For subsequent analyses, the eight genes identified here were combined with the 22 further escape genes previously characterised

(Wang et al., 2014b) to give a final set of 30 escape genes. In the following sections, I perform an analysis of allele-specific methylation on the mouse and opossum X chromosomes.

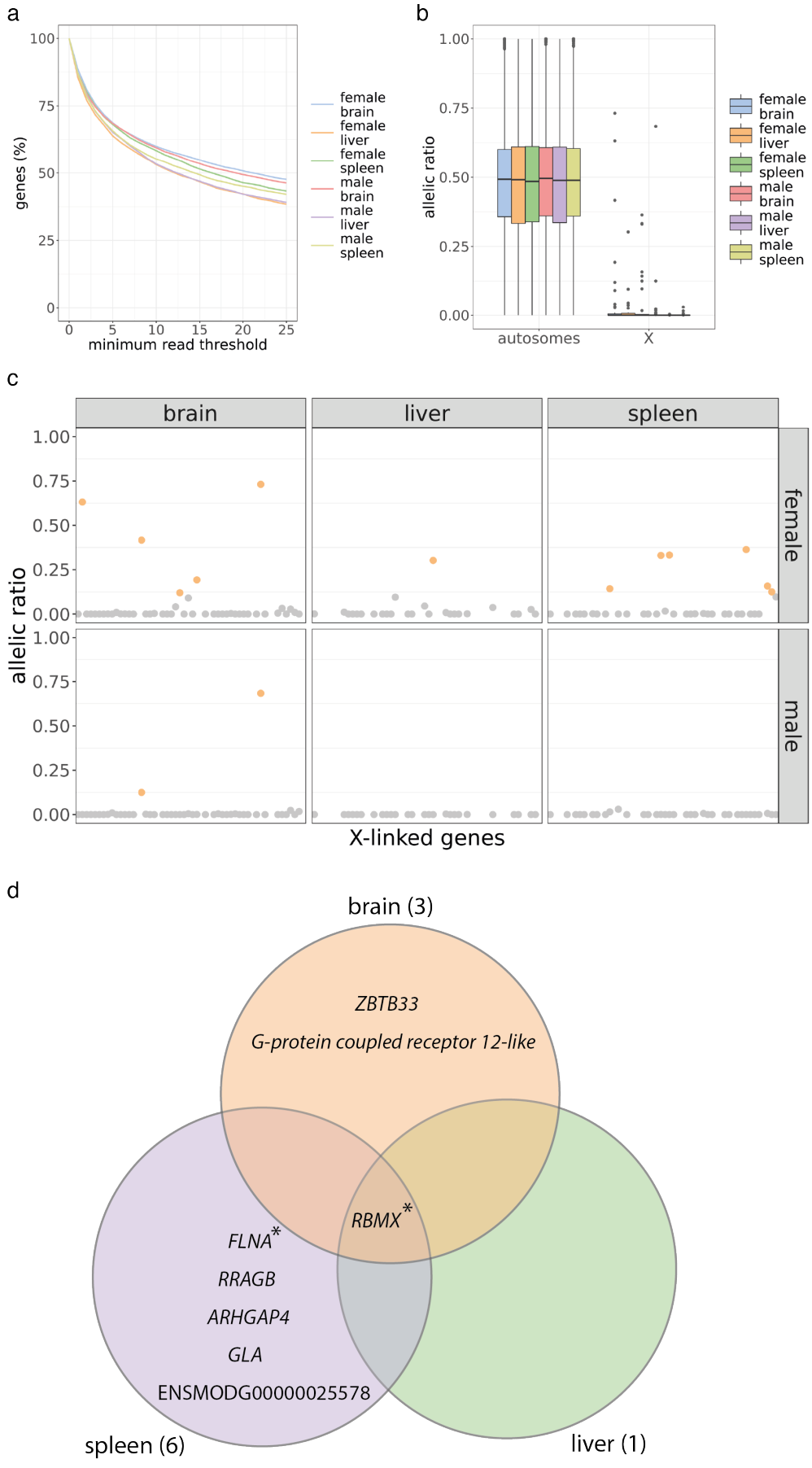

a
genes (%)
minimum read threshold
female brain
female liver
female spleen
male brain
male liver
male spleen
b
allelic ratio
autosomes
X
c
brain
liver
spleen
female
male
allelic ratio
X-linked genes
d
brain (3)
ZBTB33
G-protein coupled receptor 12-like
RBMX*
FLNA*
RRAGB
ARHGAP4
GLA
ENSMODG00000025578
spleen (6)
liver (1)

Figure 28: Allele-specific gene expression in opossum. a: Line plots showing the percentage of genes with FPKM > 1 retained for analysis under different minimum read thresholds. **b-c**: Allelic expression ratios of genes with FPKM > 1 and >9 allele-specific reads in females. Allelic expression ratios were calculated as $reads_{paternal}/(reads_{paternal} + reads_{maternal})$. $n_{brain} = 3537$; $n_{liver} = 2232$; $n_{spleen} = 2497$. **b**: Boxplots of allelic expression ratios for the autosomes and X chromosome. **c**: Allelic expression ratios of X-linked genes. Orange-coloured points indicate genes with an allelic expression ratio >0.1. $n_{brain} = 46$ $n_{liver} = 26$; $n_{spleen} = 35$. **d**: Venn diagram depicting escape genes detected in brain, liver and spleen. **RBMX* and *FLNA* were previously reported to escape XCI (Wang et al., 2014b).

4.3 Allele-specific DNA methylation profile of the mouse X chromosome

In Chapter 3 I observed that the X chromosome in female mice (the average of the Xa and the Xi) had a slightly lower methylation level than the X in males, suggesting the Xi was slightly hypomethylated. The exception to this trend was the notable hypermethylation of X chromosome CGIs in females. I also observed that in males, the X chromosome had a slightly higher methylation level than autosomes. In the present chapter, I aimed to confirm these interpretations by analysing allele-specific methylation in mouse brain and liver samples. In doing so, my goal was also to validate the analysis approach for subsequent application to opossum data.

The methylation profile of the autosomes was similar between males and females, and between maternal and paternal alleles within each sex. On average, the methylation level of autosomes was between 77 – 78% in brain and 76.3 -77.3% in liver (Figure 29a). Smoothed CpG methylation levels across the entirety of a representative chromosome (chromosome 19) demonstrated the similarity of autosomal methylation levels between sexes and parental allele-of-origin (Figure 29b).

In both tissues profiled, the female and male Xa were slightly hypomethylated relative to the autosomes (p=3.874e-4, p=1.334e-4, p=2.841e-8, p < 2.2e-16, comparison of autosomes to female Xa and male Xa in brain and liver respectively, Mann-Whitney test; Figure 29a; Table 7). The scale of the methylation difference was small, with overall methylation differences of between 1-3%. This contrasted to my findings in Chapter 3,

where I observed the X chromosome to have a higher methylation level than the autosomes in males.

Overall, the Xi was hypomethylated relative to both the female and male Xa, confirming my interpretation from Chapter 3 ($p=3.43e-5$, $p<2.2e-16$, $p<2.2e-16$, $p<2.2e-16$, comparison of the Xi to male Xa and female Xa, in brain and liver respectively, Mann-Whitney test; Figure 29a; Table 7). However, the absolute difference in methylation between Xi and Xa differed between the analyses. In the current analysis, there was a 2-3% overall methylation difference between the active and inactive X chromosome, as observed in plots of smoothed CpG methylation profiles across the X chromosome (Figure 29b). In contrast, in Chapter 3 I observed an absolute methylation difference of 7% between the male and female X chromosome.

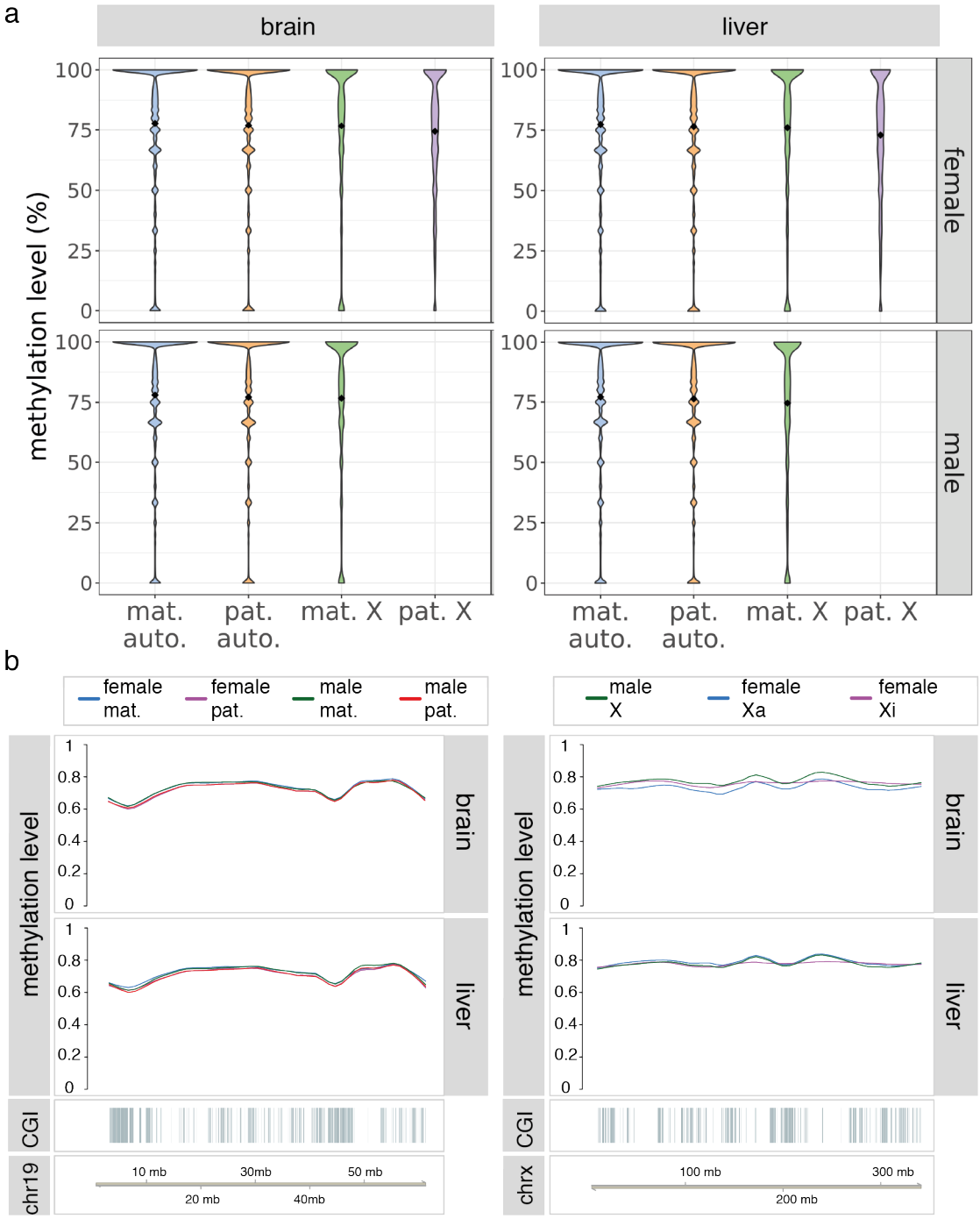


Figure 29: Genome-wide patterns of allele-specific DNA methylation in mouse. Allele-specific DNA methylation of the mouse X chromosome compared to autosomes. Methylation level is the percentage of methylated CpGs/covered CpGs within a region. CpGs sites were included if covered by a minimum of 3 reads in at least 2 replicates per condition. Data are biological triplicates. **a**: Violin plots of methylation level of 100 bp tiles in brain and liver (n_{auto} = 2,830,563; n_{chrX} = 19,827). Black diamonds indicate the mean methylation level. **b**: Genomic visualisation plot of chromosome 19 (right) and chromosome X (left). Methylation is plotted as methylation level per CpG site (n_{chr19} = 143,432, n_{chrX} = 28,489), smoothed by the 'loess' function of the R package 'Gviz'. mat. = maternal allele (C57BL/6); pat. = paternal allele (*M. spretus*); auto. = autosomes.

Brain						
Autosomes				**X chromosome**		
Female mat.	Female pat.	Male mat.	Male pat.	Female Xa	Female Xi	Male X
77.8	78	77	77.1	76.7	74.4	76.7
Liver						
Autosomes				**X chromosome**		
Female mat.	Female pat.	Male mat.	Male pat.	Female Xa	Female Xi	Male X
77.3	76.5	77.1	76.3	76	73	74.6

Table 7: Average methylation level of mouse X chromosome and autosomes (%).

To investigate methylation patterns more precisely than by whole chromosome averages, I tested individual CpG sites for differential methylation between paternal and maternal alleles in female brain and liver (sites with q-value <0.01 and a methylation difference ≥ 25%, Fisher's exact test). There was no difference in methylation between alleles on the autosomes, where 99.6% of CpG sites were unchanged between parental alleles (Figure 30a). However, allele-specific differences were apparent on the X chromosome. On the inactive X, 7.92% and 6.12% of CpG sites were hypermethylated, and 1.1% and 0.96% of CpG sites were hypomethylated, in liver and brain respectively (Figure 30a). Almost half of the hypermethylated sites were found in CGIs (45.31% in liver and 48.36% in brain). This large-scale methylation difference at CGIs was apparent in violin plots (Figure 30b). In contrast, CpGs within genes, intergenic regions, and repeat elements were highly methylated across active and inactive X chromosomes. The chromosomal average methylation levels suggested a slight hypomethylation of the Xi relative to the Xa (Table 7). To understand which features of the X chromosome might contribute to this difference, I examined methylation distributions at different genomic features (Figure 30b). A subtle difference in methylation distribution was apparent at gene bodies, where the Xi displayed fewer lowly-methylated sites, as well as a slight downwards shift among highly methylated sites (Figure 30b). Intergenic regions of the Xi also displayed a subtle downwards shift relative to the Xa (Figure 30b). These observations suggest that Xi hypomethylation cannot be attributed to a difference at one particular genomic feature (Figure 30b).

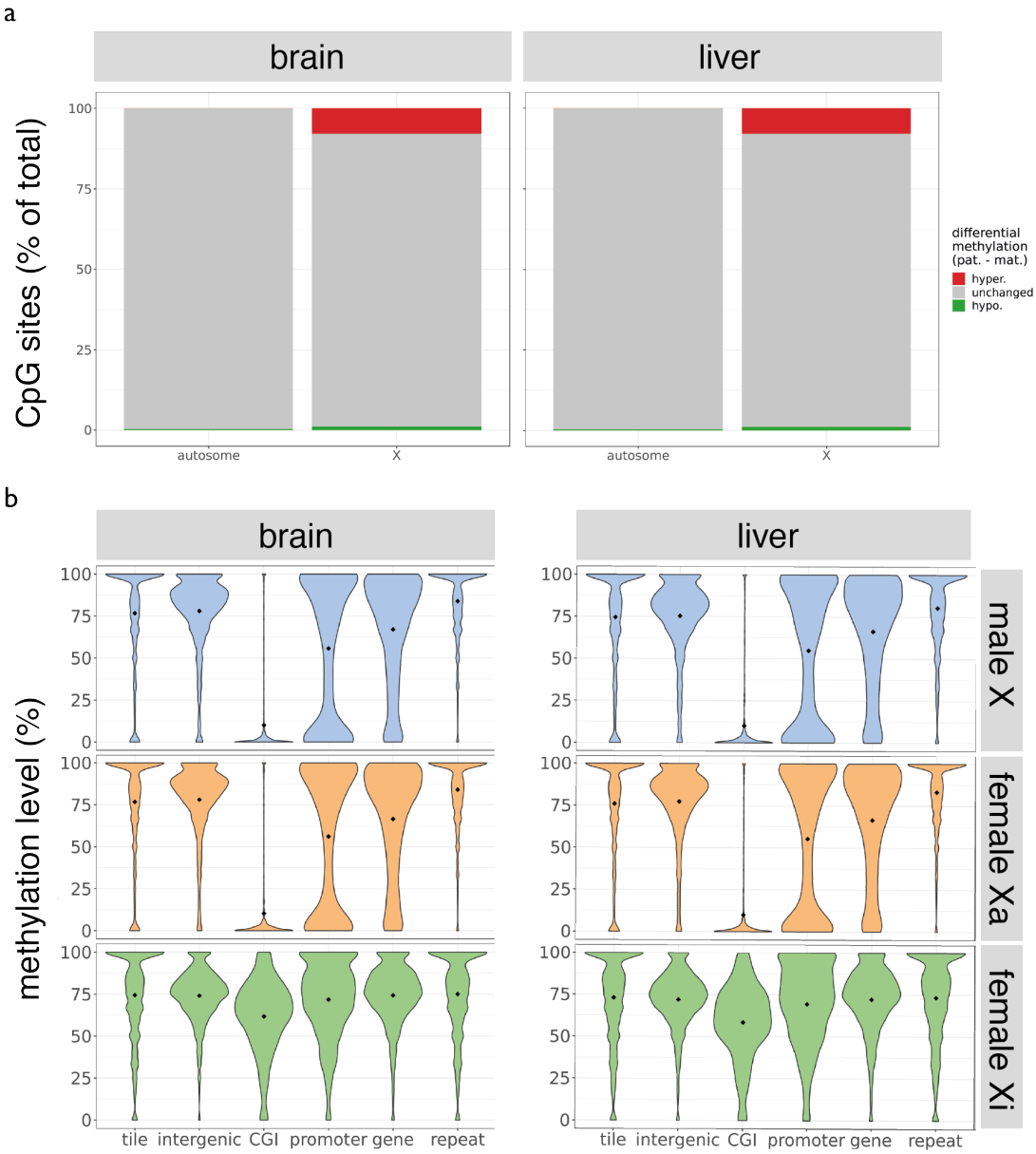


Figure 30: Allele-specific methylation of the mouse X chromosome. Methylation profile of active and inactive X chromosomes in mouse brain and liver. **a**: Percentage of CpG sites found to be differentially methylated between maternal and paternal chromosomes in female mice (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test). Red: hypermethylated on the paternal allele, green: hypomethylated on the paternal allele, grey: unchanged between the alleles. $n_{autosomal} = 1451785$, $n_{chrX} = 28489$. **b**: Violin plots depicting the methylation level of genomic features on the male X, female Xa and female Xi. Methylation level is the percentage of methylated CpGs/ covered CpGs within a region. CpGs sites were included if covered by a minimum of 3 reads in at least 2 replicates per condition. Data are biological triplicates. 100 bp tiles; non-overlapping 100 bp tiles; CGI; UCSC defined CpG islands; Promoter; Ensembl defined TSS – 2kb/+ 200 bp; Gene; Ensembl defined genes from TSS to transcription end site; Intergenic; all genomic regions not included in 'gene' category; Repeat; RepeatMasker defined repetitive elements. $n_{tiles} = 19827$; $n_{intergenic} = 1413$; $n_{CGI} = 263$; $n_{promoter} = 859$; $n_{gene} = 1093$; $n_{repeat} = 7947$.

I correlated the gene expression and methylation profiles of Xa and Xi alleles using the gene expression data generated in section 4.2.1. I plotted the methylation distribution of genes expressed in each tissue (Figure 31a), and their promoters (Figure 31b) on the male X, the female Xa, and the female Xi, plotting escape genes on the Xi separately.
Gene body methylation was similar between the male X and the female Xa (Figure 31a). Inactive genes on the Xi harboured a higher methylation level than seen on the Xa (Figure 31a). No methylation difference was observed at escape gene bodies compared to other genes on the Xi (Figure 31a). At promoters, escape genes harboured low methylation levels similar to Xa genes, while Xi genes subject to XCI showed high methylation levels (Figure 31b). Examination of methylation at exemplar loci confirmed the patterns observed overall. Methylation patterns of *Hprt,* a gene subject to XCI, demonstrated the Xi-specific hypermethylation of the CGI promoter (Figure 31c). In contrast, the CGI associated with the escape gene *Ddx3x* was not hypermethylated on the Xi although interestingly one intronic CpG was observed to carry Xi-specific methylation (Figure 31d).

The allele-specific analysis of X chromosome DNA methylation undertaken here confirmed that at a chromosome-wide level, the mouse Xi is slightly hypomethylated relative to the Xa (Figure 29). The well-reported phenomenon of Xi CGI hypermethylation was replicated by these data, and in combination with Xa and Xi specific gene expression, the Xa-like methylation pattern at promoters of escape genes was confirmed. However, I did not find evidence for the active X being more highly methylated than autosomes. Overall, these findings from allele-specific analysis verify the interpretations made based on the average X-chromosome methylation signal in Chapter 3.

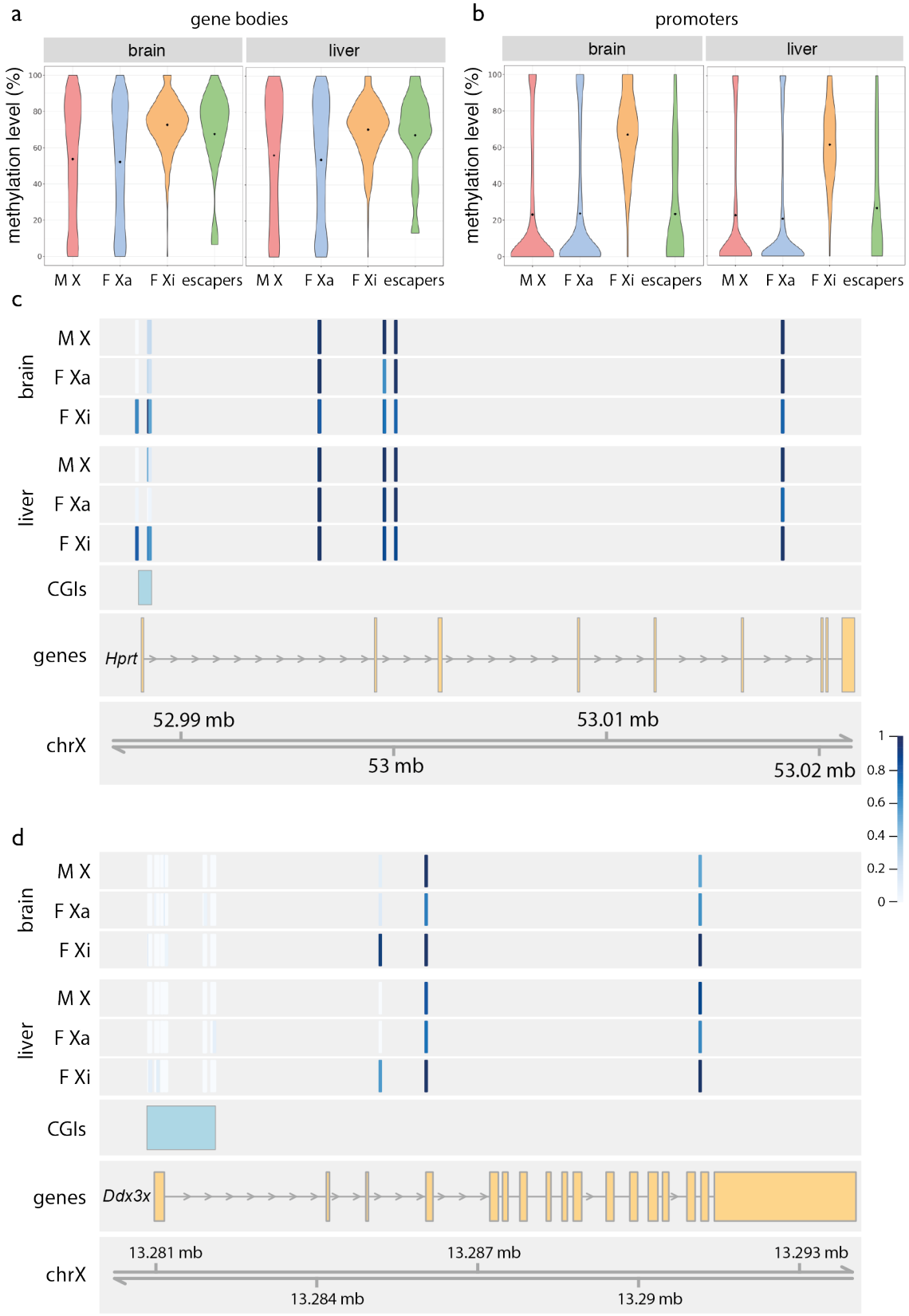

a
gene bodies
brain
liver
methylation level (%)
M X
F Xa
F Xi
escapers
b
promoters
c
brain
liver
CGIs
genes
Hprt
chrX
52.99 mb
53.01 mb
53 mb
53.02 mb
d
Ddx3x
13.281 mb
13.287 mb
13.293 mb
13.284 mb
13.29 mb

Figure 31: Methylation of mouse X-linked genes. Methylation profiles of genes and promoters on active and inactive X chromosomes in mouse brain and liver. Methylation level is the percentage of methylated CpGs/ covered CpGs within a region. CpGs sites were included if covered by a minimum of 3 reads in at least 2 replicates per condition. Data are biological triplicates. **a**: Violin plots showing methylation level of gene bodies. Genes were included if they were expressed at FPKM >1 in the respective tissue. Genes on the Xi with an allelic expression ratio of > 0.1 (Chapter 4 section 4.2) were categorised as escape genes and plotted separately from the rest of the Xi (female Xi escaper, green violin). Gene; Ensembl defined genes from TSS to transcription end site. n_{brain} = 440; n_{liver} = 320. **b**: Violin plots showing methylation level of promoters. Promoters were included following the same criteria as their associated genes. Promoter; Ensembl defined TSS – 2kb/+ 200 bp. n_{brain} = 272; n_{liver} = 194. **c**: Methylation at the *Hprt* locus, which is subject to XCI. **d**: Methylation at the XCI-escape gene *Ddx3x*.

4.4 Allele-specific DNA methylation of the opossum X chromosome

In Chapter 3, by interpreting the averaged methylation pattern of the opossum Xa and Xi, I concluded that the opossum Xi was subject to chromosome-wide hypomethylation, with the exception of escape genes. In the present analysis, I aimed to deconvolve the methylation profiles of the Xa and Xi using allele-specific analysis of gene expression and methylation in opossum brain, liver, and spleen.

Autosomal methylation levels did not differ by parental allele-of-origin, nor according to sex (Table 8, Figure 32a). Smoothed CpG methylation plots across chromosome 8 demonstrate the similarity in methylation level between the alleles (Figure 32b). In contrast to the autosomes, interesting differences in methylation were apparent when comparing the methylation of the Xa and Xi.

In all tissues, the Xi was significantly hypomethylated relative to the autosomes and the Xa ($p < 2.2e-16$ for all comparisons, Mann-Whitney test; Table 8, Figure 32a). Smoothed CpG methylation plots across the X chromosome demonstrate the marked hypomethylation of the Xi (Figure 32c). A comparison of the male X and the female Xa showed the two active chromosomes harbour very similar methylation levels (brain $p = 0.1979$, liver $p = 0.6927$, spleen $p = 0.1947$, Mann-Whitney test).

Brain						
Autosomes				**X chromosome**		
Female mat.	Female pat.	Male mat.	Male pat.	Female Xa	Female Xi	Male X
77.2	77.0	77.2	77.0	80.3	7.28	80.9
Liver						
Autosomes				**X chromosome**		
Female mat.	Female pat.	Male mat.	Male pat.	Female Xa	Female Xi	Male X
72.0	71.5	72.6	72.2	72.7	13.4	74.0
Liver						
Autosomes				**X chromosome**		
Female mat.	Female pat.	Male mat.	Male pat.	Female Xa	Female Xi	Male X
61.2	61.1	60.5	60.3	60.6	7.65	59.6

Table 8: Average methylation level of opossum X chromosome and autosomes (%).

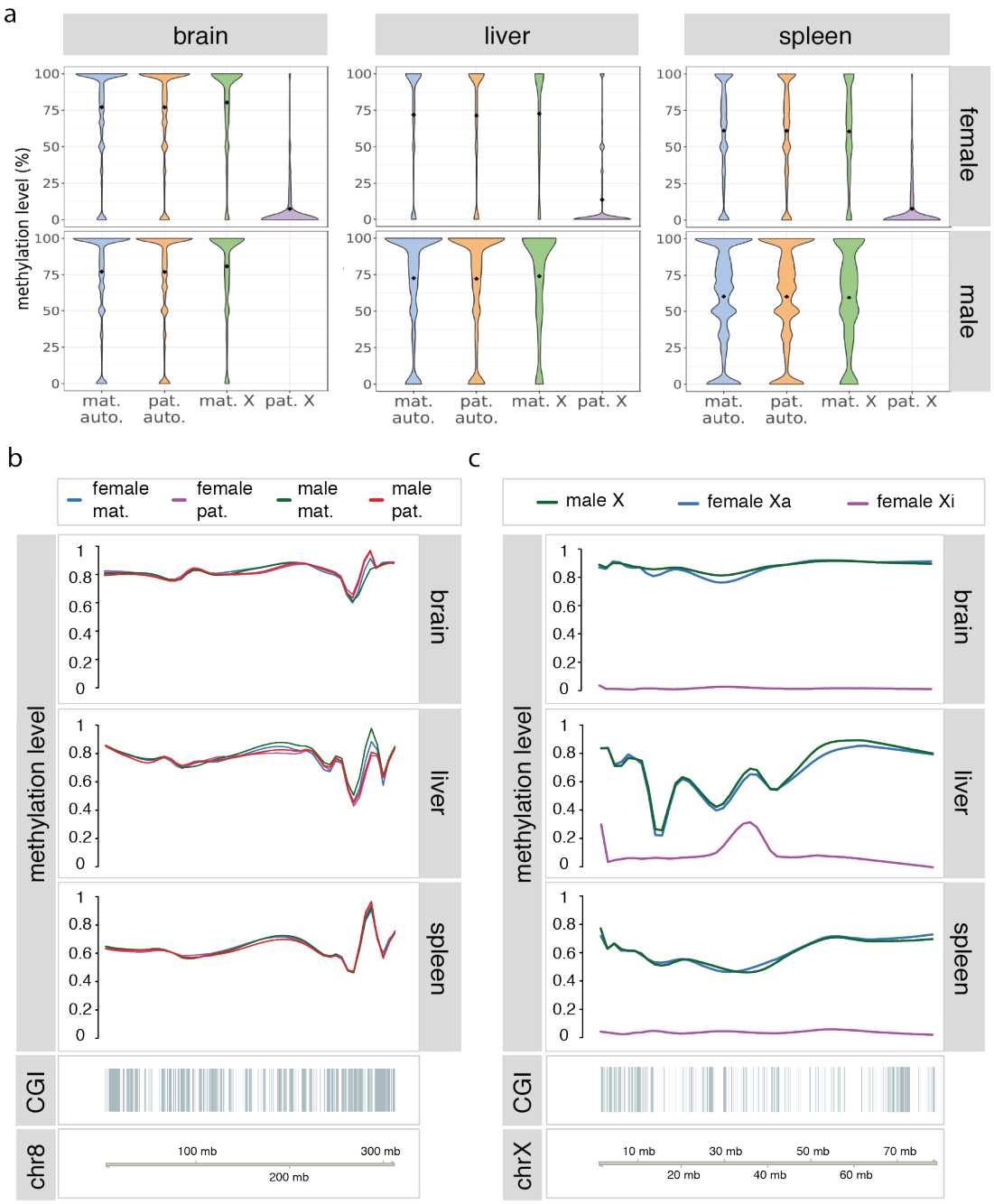


Figure 32: Genome-wide patterns of allele-specific DNA methylation in opossum. Allele-specific DNA methylation of the opossum X chromosome compared to autosomes. Methylation level is the percentage of methylated CpGs/covered CpGs within a region. CpGs sites were included if covered by a minimum of 2 reads in at least 2 replicates per condition. Data are biological triplicates. **a**: Violin plots of methylation level of 100 bp tiles in brain (n_{auto} =119388; n_{chrX} = 6063), liver (n_{auto} = 39774; n_{chrX} = 1482), and spleen (n_{auto} = 144855; n_{chrX} = 6990). Black diamonds indicate the mean methylation level. **b-c**: Genomic visualisation plots of chromosome 8 and chromosome X. Methylation is plotted as methylation level per CpG site in brain (n_{chr8} = 6667; n_{chrX} = 3365), liver (n_{chr8} = 2311; n_{chrX} = 825), and spleen (n_{chr8} =8034; n_{chrX} =3923) smoothed by the 'loess' function of the R package 'Gviz'. mat. = maternal allele; pat. = paternal allele; auto. = autosomes.

Comparison of methylation between the active X chromosome and autosomes revealed differences that were tissue-specific but consistent between the sexes. In brain, the active X was more highly methylated than the autosomes (female p = 2.682e-12, male: p = 1.657e-14, Mann-Whitney test; Table 8, Figure 32a). In liver, methylation levels did not differ between the autosomes and the active X chromosome (female p= 0.5389, male p = 0.1526, Mann-Whitney test; Table 8, Figure 32a). In spleen, the active X chromosomes had a slightly lower level of methylation than the autosomal average (female p= 0.01235, male p = 0.002978, Mann-Whitney test; Table 8, Figure 32a).

Direct comparison of methylation at individual CpG sites confirmed that the Xi is hypomethylated. Differentially methylated regions were detected between paternal and maternal alleles in female opossum samples (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test; Figure 33a).). On the autosomes, fewer than 0.2% of CpGs were differentially methylated (Figure 33a). On the X chromosomes, there were no sites of Xi hypermethylation in all three tissues profiled, but a large number of Xi sites were hypomethylated (29.8% in brain, 3.88% in liver, 8.74% in spleen; Figure 33a).

Examination of the methylation distribution at different genomic features demonstrated the similarity between the Xa of females and males. On the Xa of both sexes, gene bodies, intergenic regions, and repetitive elements carried methylation levels of approximately 60%, while promoters had a mixture of high and low methylation, and CGIs were lowly methylated (Figure 33b). In contrast, the Xi was hypomethylated at all genomic features examined, including repeat sequences (Figure 33b).

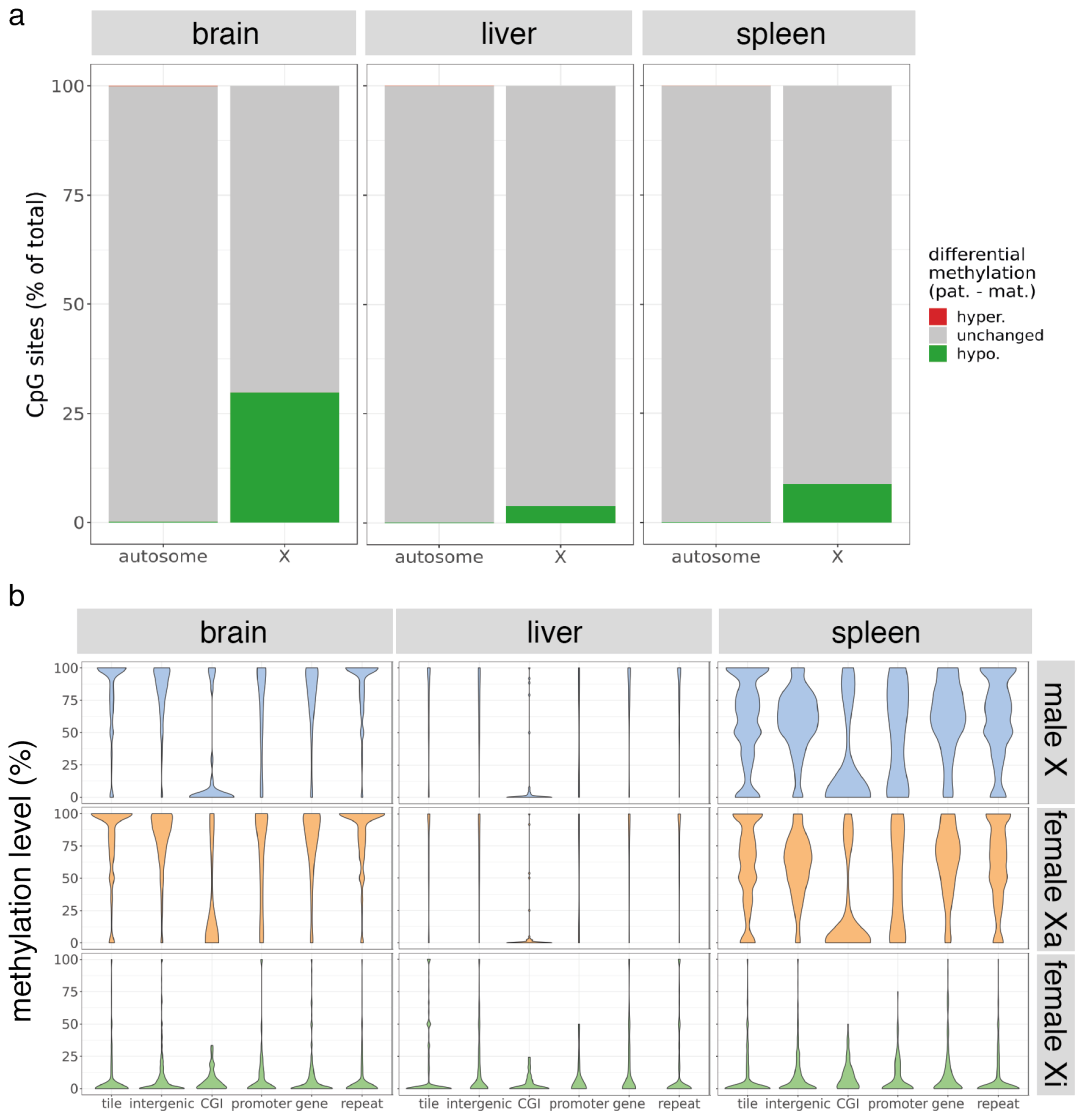


Figure 33: Allele-specific DNA methylation of the opossum X chromosome. Methylation profile of active and inactive X chromosomes in opossum brain, liver, and spleen. **a**: Percentage of CpG sites found to be differentially methylated between maternal and paternal chromosomes in female opossums (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test). Red: hypermethylated on the paternal allele, green: hypomethylated on the paternal allele, grey: unchanged between the alleles. Brain: $n_{autosomal}$ =101627 , n_{chrX} = 3365. Liver: $n_{autosomal}$ = 30938, n_{chrX} = 825. Spleen: $n_{autosomal}$ = 121817, n_{chrX} = 3923. **b**: Violin plots depicting the methylation level of genomic features on the male X, female Xa and female Xi. Methylation level is the percentage of methylated CpGs/ covered CpGs within a region. CpGs sites were included if covered by a minimum of 2 reads in at least 2 replicates per condition. Data are biological triplicates. 100 bp tiles; non-overlapping 100 bp tiles; CGI; UCSC defined CpG islands; Promoter; Ensembl defined TSS – 2kb/+ 200 bp; Gene; Ensembl defined genes from TSS to transcription end site; Intergenic; all genomic regions not included in 'gene' category; Repeat; RepeatMasker defined repetitive elements. Brain: n_{tiles} = 2021; $n_{intergenic}$ = 203; n_{CGI} = 25; $n_{promoter}$ = 43 ; n_{gene} = 132; n_{repeat} =787. Liver: n_{tiles} = 494; $n_{intergenic}$ = 115; n_{CGI} = 16; $n_{promoter}$ = 18 ; n_{gene} = 72; n_{repeat} =207. Spleen: n_{tiles} = 2330; $n_{intergenic}$ = 212; n_{CGI} = 31; $n_{promoter}$ = 44 ; n_{gene} = 145; n_{repeat} = 918.

In Chapter 3, I observed that some genes on the female X chromosome had a high methylation level, similar to the equivalent locus on the male X. I hypothesised that these were genes that escaped from XCI, and demonstrated that many of the 24 previously reported opossum escape genes were included in the group of genes displaying the unique 'escaper methylation' signature.

Integrating allele-specific methylation data with the escape genes characterised in section 4.2.2 as well as the previously defined escape genes (Wang et al., 2014b), I examined methylation levels at X-linked genes. Genes on both the female and male active X chromosome were largely hypermethylated, and non-escape genes on the Xi were mostly hypomethylated (Figure 34a). However, escape genes covered in this dataset displayed methylation levels in a similar range to the active X chromosomes. Examination of the methylation pattern surrounding *FLNA,* an XCI-escape gene, exemplifies the similarity between Xa and Xi at escape genes (Figure 34b). The *TMEM164* locus, which is subject to XCI, demonstrates the hypomethylation of the Xi (Figure 34c). This confirmed the conclusions I drew in Chapter 3, that escape genes display a methylation signature akin to the active X chromosome.

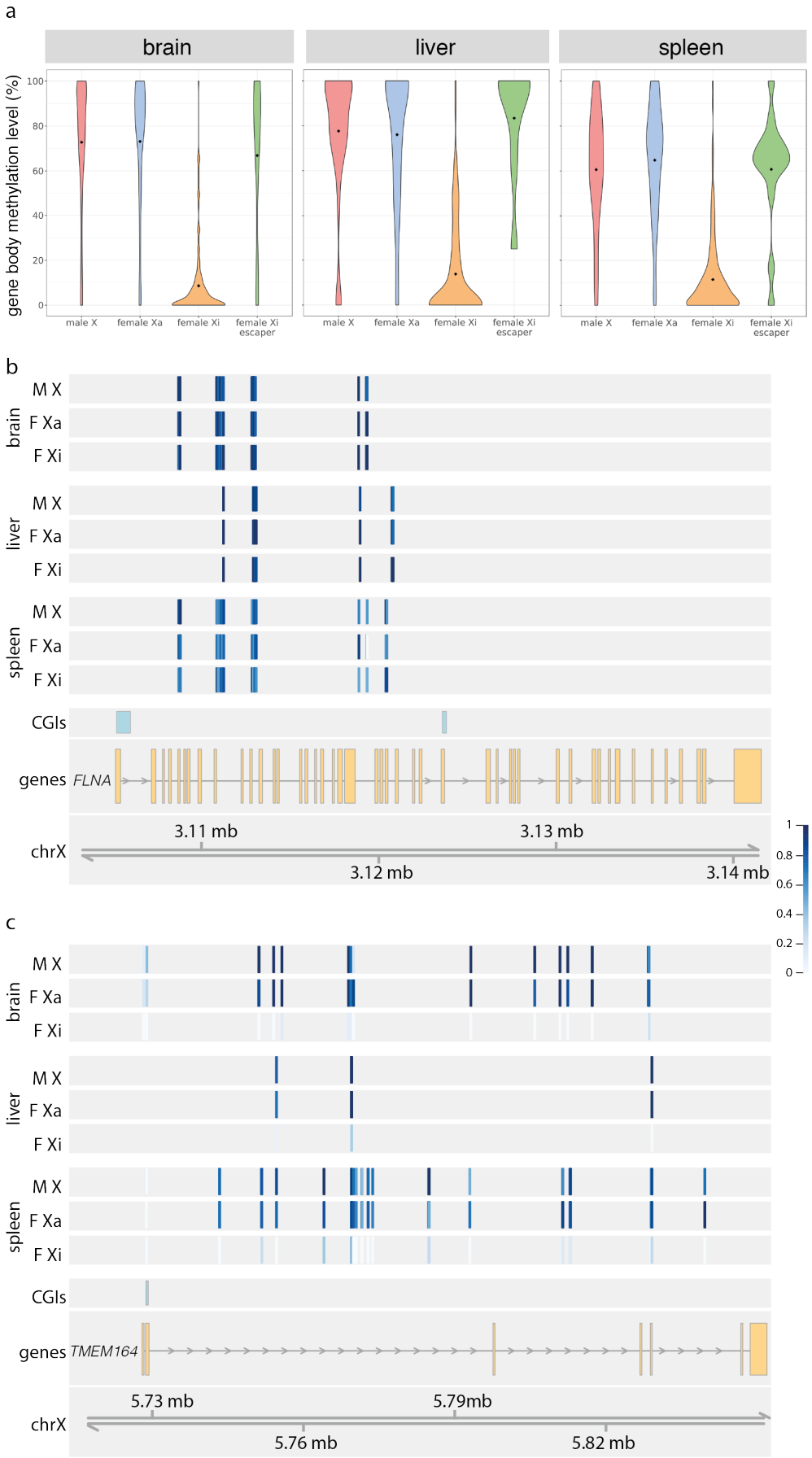

a
brain
liver
spleen
gene body methylation level (%)
100
80
60
40
20
0
male X
female Xa
female Xi
female Xi escaper
b
M X
F Xa
F Xi
brain
liver
spleen
CGIs
genes
FLNA
chrX
3.11 mb
3.13 mb
3.12 mb
3.14 mb
1
0.8
0.6
0.4
0.2
0
c
TMEM164
5.73 mb
5.79mb
5.76 mb
5.82 mb

Figure 34: Methylation of opossum escape genes. Methylation distribution at opossum X-linked genes. **a**:Violin plots showing methylation profiles of genes on active and inactive X chromosomes in opossum brain, liver, and spleen. Methylation level is the percentage of methylated CpGs/covered CpGs within a region. CpGs sites were included if covered by a minimum of 2 reads in at least 2 replicates per condition. Data are biological triplicates. Genes were included if they were expressed at FPKM >1 in the respective tissue. Genes on the Xi with an allelic expression ratio of > 0.1 (Chapter 4 section 4.2), or previously annotated as escapers (Wang et al., 2014b) were categorised as escape genes and plotted separately from the rest of the Xi (female Xi escaper, green dots). Gene; Ensembl defined genes from TSS to transcription end site. $n_{brain} = 67$; $n_{liver} = 41$; $n_{spleen} = 59$. **b**: Methylation patterns at the XCI-escape gene *FLNA*. **c**: Methylation patterns at the *TMEM164* gene, which is subject to XCI.

4.5 Discussion

In this chapter, I analysed RNA-seq and BS-seq datasets in an allele-specific manner, to profile the methylation state of mouse and opossum Xa and Xi individually, including an analysis of genes that escape XCI.

My analysis of allele-specific gene expression on the mouse X chromosome reproduced many previously reported escape genes (Berletch et al., 2015). The number of escape genes detected in my analysis was lower than in Berletch et al (2015). This is likely due to differences in analysis approach; I required both a minimum FPKM and a minimum number of allele-assigned reads to include a gene for analysis, while Berletch et al applied a binomial model that allowed them to analyse even lowly-expressed genes.

One of the tissues I profiled, the liver, had not been examined in previous surveys of mouse escape genes. Analysis of liver revealed that two genes, *Cfp* and *Tmem29*, which had previously been reported to escape XCI in only one tissue, also escaped XCI in the liver, thereby extending the list of 'common' escape genes. It is therefore possible that as a wider range of tissues are studied, other genes currently identified as tissue-specific escapers may be reassessed as 'common' escapers.

Methylation profiles of the deconvolved mouse Xa and Xi reproduced the well-reported phenomenon of Xi CGI hypermethylation, and also confirmed that the Xi is slightly hypomethylated relative to the Xa. Promoters of genes escaping XCI were confirmed to

be exempt from DNA methylation, displaying a methylation pattern akin to the equivalent allele on the Xa. I concluded that the mouse Xi is highly methylated across all types of genomic feature, with the exception of escape gene promoters.

In Chapter 3, I observed that in males, the X chromosome was slightly hypermethylated relative to the autosomes, and speculated that an XCU-related increase in transcription from the Xa might explain this increased methylation level. Changes in epigenetic marks have been associated with XCU; for example, increased levels of the active histone mark H4K16ac and the active histone variant H2AZac are observed at the promoters of Xa genes (Deng et al., 2013). In contrast to my observation in Chapter 3, in the present analysis I did not find evidence for the active X chromosome having a higher methylation level than autosomes; rather, both the male and female Xa had an average chromosome-wide methylation level the same or slightly lower than the autosomal average.

It is possible that the disparity between the findings in Chapter 3 and the current chapter derive from technical differences; different bisulfite sequencing techniques were used in the two experiments. The RRBS performed in Chapter 3 enriched for CGIs and genes, while the BS-seq performed in this chapter covered the genome in an unbiased manner. An XCU-related DNA methylation signal might be expected to occur largely at gene bodies, if it is associated with active transcription. When analysing XCU in RNA-seq data, only a subset of X-linked genes should be considered, excluding those genes with specialised expression patterns (Deng et al., 2011, Deng et al., 2013, Sangrithi et al., 2017). Future work to investigate a potential link between Xa methylation level and XCU in more detail could focus on this refined gene list.

While the mouse allele-specific analysis was facilitated by the well-established C57BL/6J x *M. spretus* system, a similar approach in opossum required the cataloguing of genetic variants between the parents of an opossum cross. Far fewer SNPs were available between the opossum parents than for the mouse cross, which was expected from a cross of closely related animals.

For assignment of allele-specific reads, I relied on a publicly available tool, SNPsplit, which exploited only unambiguous SNPs; i.e. those SNPS that were homozygous on the maternal X, and hemizygous for an alternative allele on the paternal X. Future work could develop a tailored approach to recover information from other X-linked positions that could in principle be informative; for example at positions where the mother is heterozygous and one maternal allele is shared with the paternal X chromosome. In this case, if the daughter is heterozygous, reads at this position could be assigned to the parent of origin. Such an approach would increase the number of X-linked sites available for analysis. ChrRNA-seq, a method which captures intronic as well as exonic transcripts, could also be employed in future as a means of increasing SNP coverage (Gdula et al., 2019, Nojima et al., 2015).

While the relative paucity of informative SNP sites meant that far fewer loci were able to be analysed in allele-specific manner on the opossum X chromosome than the mouse, for the sites I was able to profile, my results confirmed the conclusions drawn in Chapter 3. RRBS experiments in Chapter 3 showed that the X chromosome was hypomethylated in females compared to males, supporting previous work (Waters et al., 2018). From this data I drew the conclusion that the opossum Xi was hypomethylated, based on the assumption that the Xa in females would have an equivalent methylation level to the male X. The allele-specific analysis in this chapter confirmed this, revealing the female Xa and male X chromosome to have remarkably similar levels of methylation. By contrast, the opossum Xi was revealed to be globally hypomethylated relative to both the Xa and the autosomes in tissues representing all three developmental germ layers. The hypomethylation extended across all genomic features, including genes, promoters, CGIs and repetitive elements.

I performed an analysis of allele-specific gene expression on the opossum X chromosome, aiming to identify genes that escape XCI. This analysis identified two genes previously reported to escape XCI, and uncovered six novel escape genes. The number of escape genes detected in this analysis was lower than in a previous survey (Wang et al., 2014b). This is unlikely to be due to differences in analysis approach; both the application of an FPKM threshold, and the use of an allelic ratio of > 0.1 were also

employed in Wang et al. (2014b). It is possible that a higher number of escape genes would have been detected in the opossum if a less stringent lower threshold for number of assigned allelic reads had been applied. However, as allelic ratios were 'noisy' at loci with very low numbers of allele-assigned reads, it was necessary to be conservative in the interpretation of these data to avoid false positives. The lower number of escape genes detected in this analysis is more likely due to differences between the opossum laboratory stocks used in this work and the previous survey.

Two of the six newly identified opossum escape genes, two genes (*RRAGB* and *GLA*) were identified as putative escape genes based on their methylation profile in Chapter 3. Of the other four novel escape genes, three were not covered in the Chapter 3 analysis, and one, *ARHGAP4*, had brain and liver methylation ratios below 0.8. However, *ARHGAP4* was identified here as tissue-specific escape gene in the spleen, a tissue which was not profiled in Chapter 3. These findings supported my conclusion in Chapter 3 that genes for which female and male methylation levels are similar should be considered putative escape genes and identified two cases in which this was unequivocally true. In further support of my conclusion, examination of allele-specific gene body methylation confirmed that escape gene bodies had methylation levels equivalent to active X genes.

If all genes identified as putative escapers based on their methylation profile are similarly *bona fide* escapers, this would mean approximately 15% of opossum X chromosome genes escape XCI. Ideas of metatherian XCI as 'leaky' are prevalent in the literature (Cooper et al., 1993, Samollow et al., 1989, Migeon et al., 1989). However, recent analyses of metatherian XCI rejected the idea of 'leaky' expression (Mahadevaiah et al., 2009, Julien et al., 2012). Between 3-7% of mouse genes escape XCI (Berletch et al., 2015), but up to 8-15% of X-linked genes escape XCI in humans (Carrel and Willard, 2005, Zhang et al., 2013b, Cotton et al., 2013). Therefore, a value of up to 15% of opossum genes escaping XCI is similar to what has been reported for humans, and in fact the mouse may be exceptional compared to other mammals (Berletch et al., 2015).

In Chapter 3, RRBS experiments found the active X in male opossum to have a higher average methylation level than autosomes in both brain and liver. However, in the present

BS-seq analysis, I found different trends depending on the tissue. In the brain the Xa and the male X chromosome had a higher average methylation level than the autosomes, while in liver there was no difference. In spleen the autosomes were slightly but significantly more methylated than the active X chromosomes. This suggests that methylation level of the X chromosome may relate to specific gene expression/regulation requirements of each tissue.

Unlike the RRBS technique used in Chapter 3, BS-seq libraries capture sites across the genome in an unbiased manner. In this chapter I examined methylation level at those X-linked repetitive elements that were captured in the allele-specific analysis. I found that repetitive elements on the Xi were hypomethylated. This intriguing finding raises interesting questions about the silencing of repeat elements on the opossum X chromosome. In future work it would be interesting to establish whether repetitive elements on the opossum Xi are at increased risk of de-repression. Epigenetic mechanisms other than DNA methylation are implicated in the control of repetitive elements. Perhaps alternate mechanisms are responsible for the control of repetitive elements on the opossum Xi, as is presumably also true in the case of promoter silencing, owing to the lack of CGI DNA methylation.

It is interesting to consider at what point in development the hypomethylated state of the Xi originates. In eutherians, the genome passes through two phases of global hypomethylation; once in the germ cell, and once in the early preimplantation embryo. The final methylation state of the Xi, especially including CGIs, is therefore established during the wave of remethylation in the peri- implantation embryo. In metatherians, a similar genome-wide reprogramming phenomenon occurs in the germline (Suzuki et al., 2013, Ishihara et al., 2019), but the methylation state of the early embryo has not yet been reported. In Chapter 5, to address questions regarding X-chromosome methylation status during embryogenesis, as well as more general questions concerning the nature of genome-wide reprogramming in metatherians, I perform methylome profiling of opossum gametes and early embryos.

Chapter 5. Results 3

Eutherians harbour high levels of methylation across most of the genome, with the exception of CpG islands. While the methylation pattern of differentiated tissues is relatively static, during other phases of the life cycle, methylation levels are subject to vast dynamic changes associated with the reprogramming of the genome. This first occurs in primordial germ cells (PGCs). In the early embryo, as PGCs are specified and begin to migrate to the genital ridge, DNA methylation is erased from the genome (Seki et al., 2005, Hajkova et al., 2010, Popp et al., 2010, Guibert et al., 2012, Seisenberger et al., 2012). Specific methylation patterns are then re-established in oocyte and sperm during gametogenesis (Seisenberger et al., 2012, Shirane et al., 2013, Singh et al., 2013, Veselovska et al., 2015). This reprogramming is thought to be necessary for the erasure of epigenetic memories, and includes the removal of parental imprints and reactivation of the inactive X chromosome: essentially a 'resetting' of the genome.

A second reprograming event is initiated in eutherian zygotes. The highly methylated sperm nucleus is rapidly demethylated, to levels almost as low as the oocyte, in a DNA replication-independent manner, confirming a phenomenon that had been previously observed by immunofluorescence of early embryos (Oswald et al., 2000, Mayer et al., 2000, Smith et al., 2012, Smith et al., 2014, Guo et al., 2014b). Both parental genomes then undergo further demethylation by a passive mechanism, reaching a low point in the inner cell mass of the pluripotent blastocyst. The activation of the embryonic genome (EGA) therefore occurs during a phase of unique hypomethylation in eutherian mammals. Remethylation of the embryonic genome is achieved via the action of *de novo* methyltransferases at the exit from pluripotency, and methylation patterns take on a genome-wide profile broadly resembling adult tissues.

The pre-implantation phase of hypomethylation is unique to mammals, as similar profiling experiments performed in zebrafish showed no sustained hypomethylation phase. After fertilisation, the zebrafish oocyte and sperm genomes undergo localised changes that produce, by the time of EGA, a genome methylation pattern resemblant of differentiated tissues (Potok et al., 2013, Jiang et al., 2013). The zebrafish germline also

does not undergo a genome-wide reprogramming (Ortega-Recalde et al., 2019, Skvortsova et al., 2019).

Several hypotheses have been suggested to explain the eutherian requirement for reprogramming. One suggestion is that the rapid demethylation of the paternal genome has evolved as a maternal defence system against imprints carried by the sperm, which is consistent with the conflict theory of imprinting (Reik and Walter, 2001b). A second view states that demethylation permits the early and widespread activation of transcription, for which there is a unique requirement in eutherian embryos because of the specification of the trophoblast lineage (Hackett and Surani, 2013). This notion is supported by an experiment in which depletion of DNMT1 in *Xenopus* embryos resulted in precocious embryonic genome activation (Stancheva and Meehan, 2000). A final hypothesis suggests that reprogramming is essential for the removal of specialized gametic epigenetic marks, in order to produce a totipotent zygotic genome (Hackett and Surani, 2013, Monk et al., 1987).

Studies using immunofluorescence and locus-specific bisulfite profiling of DNA methylation reported that genome-wide reprogramming occurs in the marsupial germline (Suzuki et al., 2013, Ishihara et al., 2019). This fits with the knowledge that marsupials possess imprinted genes, at least one of which is suggested to rely on a differentially methylated region for its regulation (Smits et al., 2008). Given that a suggested function for reprogramming of methylation in the germline is imprint resetting, a requirement for germline reprogramming would be expected in the marsupial lineage. However, whether or not there is a global reprogramming of DNA methylation in the metatherian early embryo remains unknown. Marsupials possess the key mammalian features that have been suggested to cause the requirement for global demethylation in eutherians; namely, marsupial embryos must also specify an extra-embryonic lineage, and the activation of the embryonic genome correspondingly occurs early in development, at the E3.5-4.5 stage (Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished). Resolving the question of whether genome-wide reprogramming occurs in marsupial embryos could shed light on the evolutionary reasons for the evolution of this phenomenon in mammalian embryos.

X-chromosome inactivation is mediated in eutherians by the long non-coding RNA *Xist* (Brockdorff et al., 1991, Borsani et al., 1991, Brown et al., 1991a, Penny et al., 1996, Marahrens et al., 1997). *Xist* is transcribed from the future inactive X, which is chosen at random in the early embryo, and spreads along the chromosome in *cis.* Following *Xist* expression, repressive epigenetic marks are recruited to the chromosome, establishing a silent state (Plath et al., 2003). High methylation levels at promoters of inactive X (Xi) genes are required for maintenance of the inactive state (Cotton et al., 2011, Sado et al., 2000, Norris et al., 1991). Establishment of CGI methylation on the inactive X chromosome occurs during the global remethylation phase associated with the exit from pluripotency, and is mediated by the *de novo* methyltransferase Dnmt3b, in concert with the structural maintenance of chromosome protein Smchd1 (Blewitt et al., 2008, Gendrel et al., 2012).

Unlike eutherians, metatherians do not rely on *XIST* for XCI. In fact, *XIST* is not found in metatherian genomes, and *LNX3,* the gene that evolved to become *XIST* in the eutherian lineage, retains protein-coding function in metatherians (Duret et al., 2006, Hore et al., 2007, Shevchenko et al., 2007). In metatherians, a lncRNA termed *RSX* is presumed to enact silencing of the inactive X chromosome (Grant et al., 2012). Metatherian XCI is imprinted, such that *RSX* is always expressed from, and silences, the paternal X chromosome (Grant et al., 2012). The molecular nature of this imprint is not known, but a region of the *RSX* promoter was shown to be highly methylated in the foetal brain and extra-embryonic tissues of males and intermediately methylated in females, suggesting that differential methylation of the *RSX* promoter between the Xa and Xi might mediate imprinted expression (Wang et al., 2014b). However, it remains to be established whether this pattern originates in germ cells as a true germline DMR. The recent discovery of *XSR*, an antisense transcript at the *RSX* locus, suggests another possibility. *XSR* is expressed in oocytes, and from the active X in embryos, and its transcription unit entirely encompasses *RSX* (Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished). *XSR* transcription from the active X could prevent *RSX* expression in *cis* as a means of regulating monallelic *RSX* expression, in a manner similar to loci with reciprocally imprinted lncRNAs in eutherians. Therefore, a differential methylation mark

at the *XSR* promoter is also a potential candidate for the molecular nature of the XCI imprint.

In this chapter I therefore profiled DNA methylation across a time-course of opossum embryonic development encompassing gametes and embryos from E1.5 - 7.5 . I aimed to answer whether there is global DNA methylation reprogramming in the marsupial embryo, and also to examine the dynamics of X chromosome DNA methylation during XCI. I also examined the methylation state of the *RSX* locus during the early developmental time period in which XCI occurs.

5.1 Evaluation of low-input bisulfite sequencing methods for single embryos

To profile methylation dynamics during the early stages of metatherian development, it was necessary to employ a bisulfite sequencing method amenable to single embryos as input material. I therefore tested different low-input bisulfite sequencing approaches, initially using mouse embryos and dilute gDNA as optimisation samples.

5.1.1 **Low-input RRBS**

Low-input RRBS libraries were prepared following a method previously reported for inputs as low as single cells (Guo et al., 2013, Guo et al., 2015b). I tested library preparation conditions using a dilution series of purified mouse gDNA to establish the success of the protocol before applying it to embryonic samples (Figure 35a-b). Bioanalyzer analysis showed that libraries of the expected size range (~170-500 bp, from previous high-input RRBS) were successfully prepared from input as low as 0.5 ng gDNA (red arrowhead; Figure 35c). The same approach was therefore applied to individual mouse E3.5 embryos.

RRBS libraries within the expected size range were successfully prepared from individual mouse E3.5 blastocysts (Figure 35d). The library peaks were narrower (~170- 200 bp) than observed in high-input RRBS, suggesting that the insert size of library fragments was smaller in low-input libraries. There were no peaks within the expected library size range in ‘no embryo’ controls, confirming that exogenous DNA contamination was avoided during library preparation. 24 cycles of library amplification PCR yielded a higher library concentration than 20 cycles (Figure 35d, compare upper and lower panels). This library (red arrow) was therefore chosen for pilot sequencing, to assess whether the preparation method would yield libraries with adequate mapping efficiency, genome coverage, and bisulfite conversion rate.

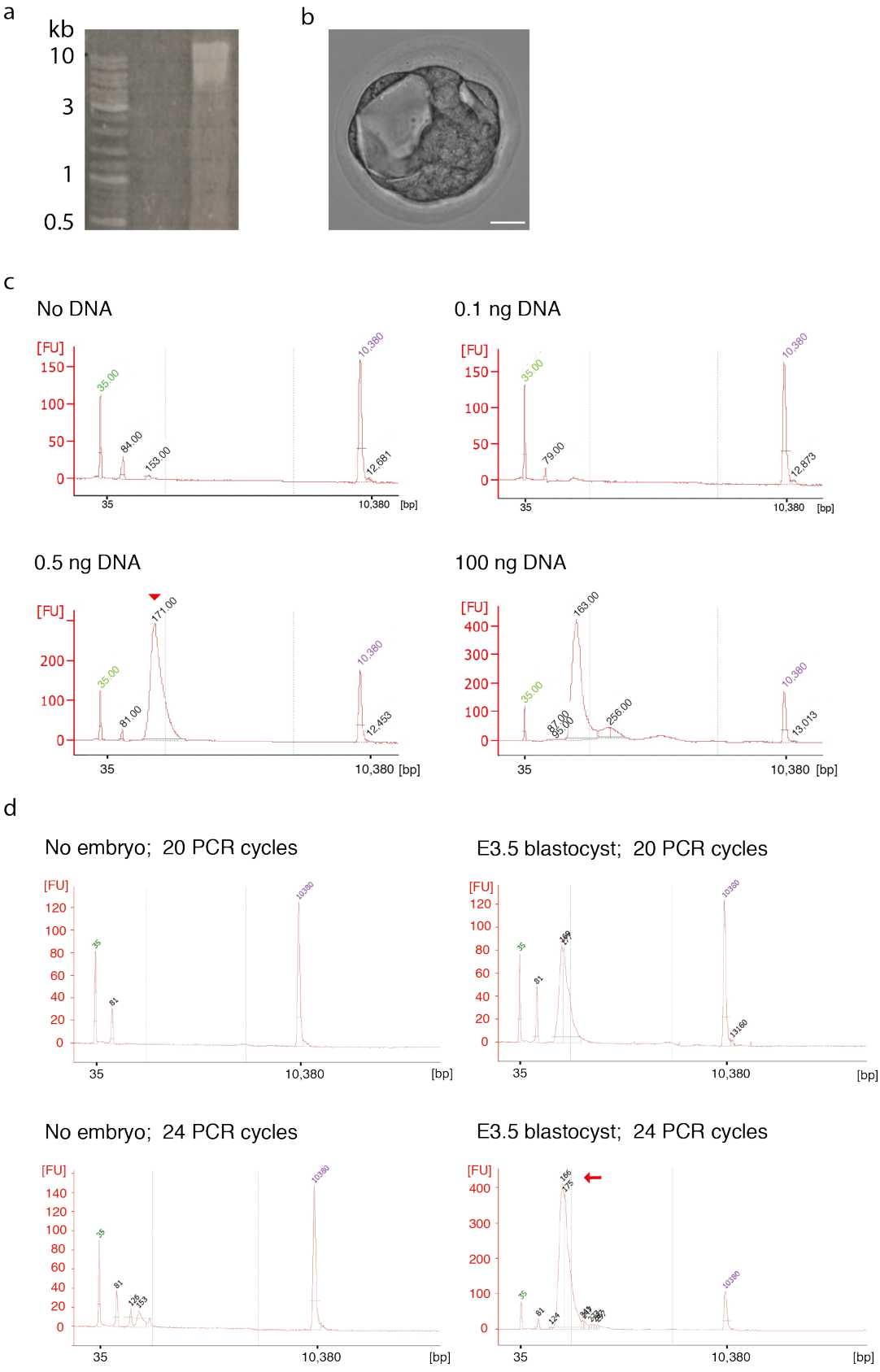

a
kb
10
3
1
0.5
b
c
No DNA
[FU]
150
100
50
0
35.00
84.00
153.00
12.681
10,380
35
10,380 [bp]
0.1 ng DNA
[FU]
150
100
50
0
35.00
79.00
12.873
10,380
35
10,380 [bp]
0.5 ng DNA
[FU]
200
100
0
35.00
81.00
171.00
12.453
10,380
35
10,380 [bp]
100 ng DNA
[FU]
400
300
200
100
0
35.00
87.00
95.00
163.00
256.00
13.013
10,380
35
10,380 [bp]
d
No embryo; 20 PCR cycles
[FU]
120
100
80
60
40
20
0
35
61
10380
35
10,380
[bp]
E3.5 blastocyst; 20 PCR cycles
[FU]
120
100
80
60
40
20
0
35
61
13160
10380
35
10,380
[bp]
No embryo; 24 PCR cycles
[FU]
140
120
100
80
60
40
20
0
35
61
126
153
10380
35
10,380
[bp]
E3.5 blastocyst; 24 PCR cycles
[FU]
400
300
200
100
0
35
61
124
166
175
10380
35
10,380
[bp]

Figure 35: Low-input RRBS library preparation optimisation. Preparation of low-input RRBS libraries. **a**: Purified genomic DNA used for optimisation libraries. **b**: Representative image of mouse E3.5 embryo used for optimisation libraries. Scale bar represents 20 microns.**c-d**: Bioanalyzer profiles of low-input RRBS libraries. Y axes show library concentration (fluorescence units; FU). X axes show library size distribution (base pairs; bp). Bioanalyzer size markers are shown in green (35 bp) and purple (10, 380 bp). **c**: optimisation trial libraries prepared from a dilution series of purified gDNA. **d**: libraries prepared from individual mouse E3.5 blastocysts (right panels), or 'no embryo controls' (picking buffer processed in parallel to control for contamination from exogenous DNA; left panels). The result of 20 (upper panels) vs 24 (lower panels) cycles of library amplification PCR is compared.

100 bp paired-end sequencing of the pilot low-input library produced approximately 30 million reads, with average sequence quality (Phred) score of greater than 30. The library was depleted of cytosines in a manner characteristic of successful bisulfite conversion (Figure 36a). However, an increase in cytosine content was apparent from approximately 50 bp onwards (Figure 36a). This increase was attributable to adapter content resulting from read-through into the 3' adapter as a result of short insert size (Figure 36b). After adapter trimming, cytosine content in the latter half of the read dropped below 10%, similar to the levels in the first half of the read (Figure 36a).

Reads were aligned to the mouse reference genome (mm10) using Bismark, with a resultant mapping efficiency of 65%. In contrast to the coverage attained in high-input RRBS, 3.8 % of CpG sites were captured in the low-input RRBS library (Figure 36c). Most captured sites were covered at a depth of at least 4x, implying that the library had been sequenced to saturation. This breadth of coverage was lower than the approximately 4.5% (~ 1 million CpGs) reported in the original single-cell RRBS protocol (Guo et al., 2015b).

Approximately 90% of mapped read ends consisted of MspI or BfaI restriction enzyme motifs (Figure 36d), suggesting most library fragments derived from the expected restriction products. It is interesting to note the difference in MspI motif frequency between this low-input library and the previous high-input RRBS libraries, in which BfaI was the predominant motif (Figure 36d).

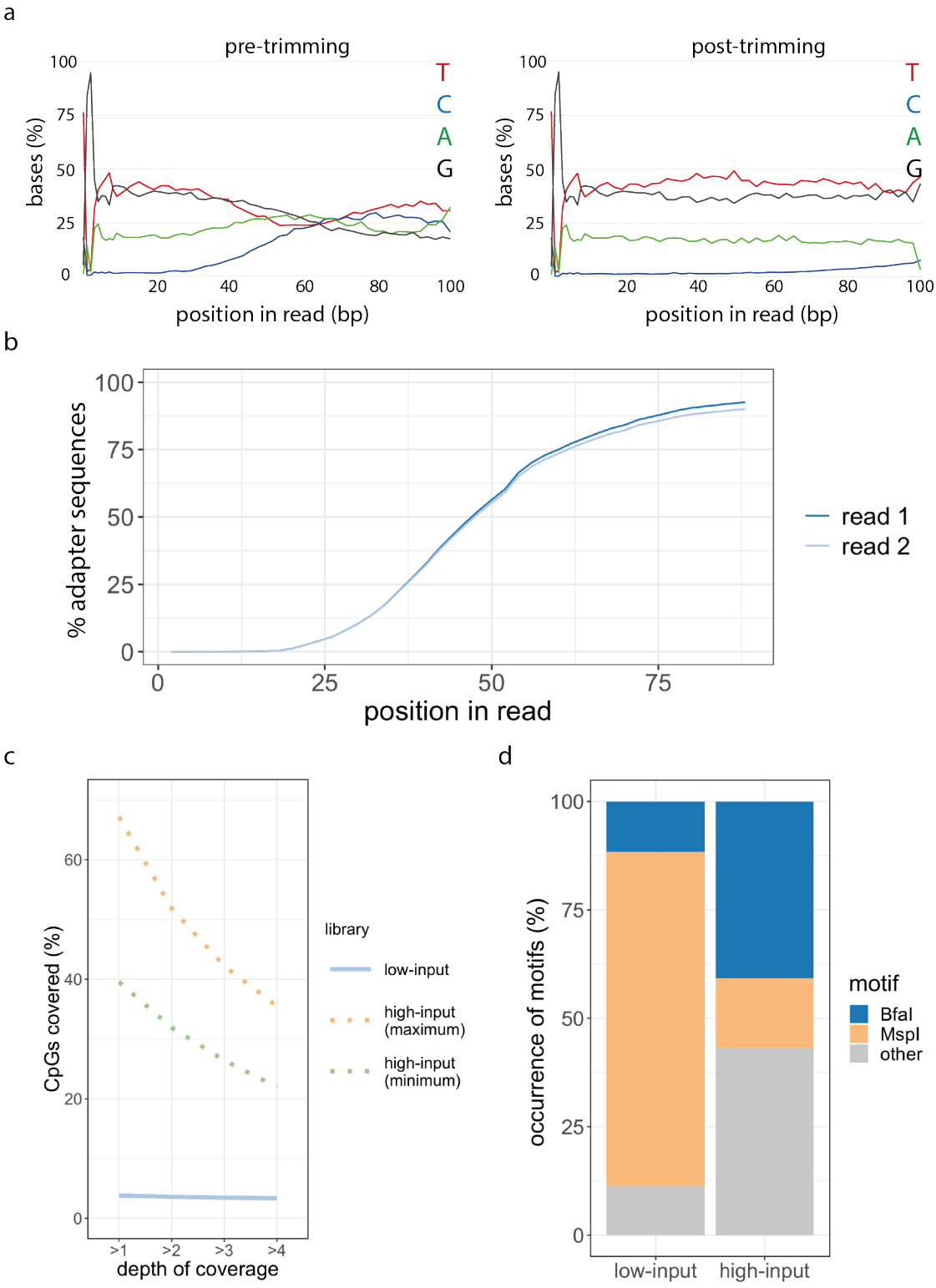

a
pre-trimming
post-trimming
bases (%)
position in read (bp)
T
C
A
G
b
% adapter sequences
position in read
read 1
read 2
c
CpGs covered (%)
depth of coverage
library
low-input
high-input (maximum)
high-input (minimum)
d
occurrence of motifs (%)
low-input
high-input
motif
BfaI
MspI
other

Figure 36: Post-sequencing quality control of the pilot low-input RRBS library. Post-sequencing assessment of the quality of the low-input RRBS library prepared from a single mouse E3.5 blastocyst. **a**: Sequence composition as a function of position in the read for read one pf the pilot library before (left panel) and after (right panel) trimming to remove adapter content. **b**: Cumulative percentage of adapter content as a function of position in the read. **c**: Percentage of CpG sites covered at different read depths. The minimum and maximum coverage of mouse high-input RRBS libraries are shown for comparison (dotted lines; data replotted from Chapter 3.2 Figure 13). **d**: Percentage of insert ends containing the restriction motifs of MspI, BfaI or any other sequence ('other') in the low-input RRBS library ('low-input', left bar). The average occurrence of restriction enzyme motifs in high-input RRBS libraries prepared from adult mouse gDNA is presented for comparison ('high-input', right bar; data replotted from Chapter 3.2 Figure 13).

The methylation patterns observed in the pilot single-embryo library recapitulated the patterns reported for the early mouse blastocyst (Smallwood et al., 2011, Smith et al., 2012, Guo et al., 2013). Comparison of methylation levels in the blastocyst and the adult male liver revealed a broadly hypomethylated genome at CGIs, gene bodies, and 5 kb genomic windows, consistent with the low methylation level reported for the pre-implantation period (Figure 37a).

Imprinted differentially methylated regions (DMRs; Schulz et al., 2008) are expected to display intermediate methylation level, reflecting the average signal of the two parental alleles. Of the 16 DMRs covered in this dataset two loci (*Nnat* and *Igf2/Air*) were 50% methylated, as expected at imprinted loci (Figure 37b). Methylation level was almost zero at the *Nesp* DMR, in line with the fact that this DMR is not imprinted in the germline or in mouse morulae (Coombes et al., 2003). Several other loci were either 0% or 100% methylated, perhaps indicating the occurrence of allelic dropout during library preparation.

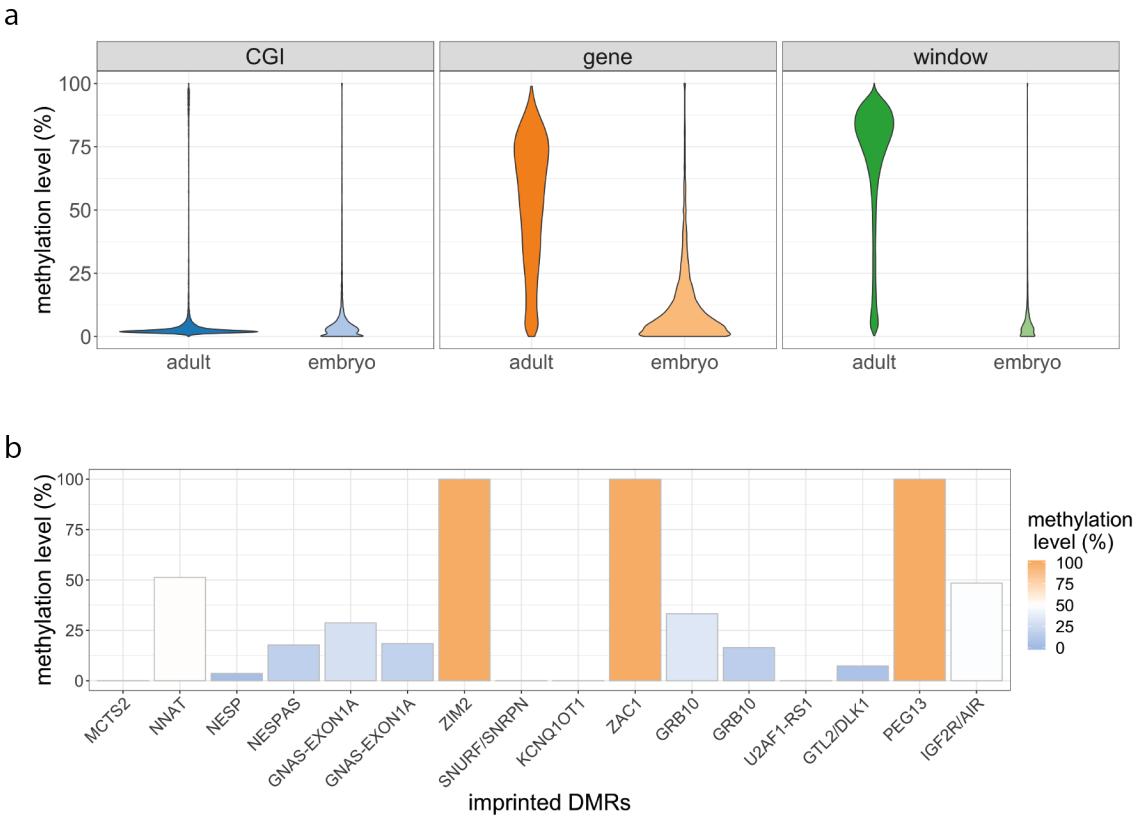


Figure 37: Methylation state of the pilot mouse blastocyst library. a: Violin plots representing the global methylation distribution in an E3.5 mouse embryo at different genomic features. Adult male liver data replotted from section 3.3 is shown for comparison. CGIs; UCSC defined CpG islands, n_{embryo} = 9852, n_{adult} = 11284. Gene bodies: Ensembl defined genes from TSS to transcription end site, n_{embryo} = 14024, n_{adult} = 21188. 5 kb windows: non-overlapping 5 kb windows across the genome, n_{embryo} = 14500, n_{adult} = 138441. **b**: Methylation level at 16 differentially methylated regions (DMRs) associated with imprinted genes. DMRs; WAMIDEX defined regions converted to mm10 coordinates using the UCSC genome browser liftOver tool.

The results of the pilot sequencing of one low-input RRBS library validated the efficacy of this method in profiling global methylation distributions, though the results at imprinted DMRs suggested that locus- or allele-specific methylation might be below the resolution of this approach. The differences in occurrence of MspI and BfaI restriction motifs between low-input and high-input RRBS libraries suggested that library preparations might be subject to batch variation in restriction digestion. In addition, the pilot low-input RRBS library showed lower than expected coverage of CpG sites, and quite high adapter content. I therefore considered whether another recently published low-input bisulfite sequencing approach would be suitable.

5.1.2 **Low-input BS-seq**

Whole-genome bisulfite sequencing from single cells (scBS-seq) can cover up to 50% of genomic CpGs and is superior to RRBS in terms of coverage of genomic features other than CGIs (Clark et al., 2016). The scBS-seq method is also amenable to 'low-bulk' samples (10s – 1000s of cells; low-input BS-seq) and presented an alternative option for profiling single opossum embryos. I therefore applied the low-input BS-seq technique to purified gDNA and single mouse embryos as pilot samples (Figure 38a-b).

Initial trials successfully amplified libraries with the expected size distribution from both gDNA and mouse blastocysts (Figure 38c). Two negative controls were employed; a water only ('no DNA') control, and an embryo picking buffer ('no embryo') control. Both controls did not amplify a library, indicating that contamination from exogenous DNA was successfully avoided during sample collection and library preparation (Figure 38c). The scBS-seq protocol uses the iPCRTag adapter system, which requires the use of a custom indexing primer for sequencing (Clark et al., 2016). To simplify the sequencing of BS-seq libraries in the Francis Crick Advanced Sequencing Facility, where such a custom primer was not standard practice, I adopted a modified oligonucleotide design that permitted the use of the NEBNext adapter system (Anna Leichter, personal communication). Libraries were successfully prepared from both gDNA and mouse blastocysts using this method, while negative controls did not amplify (Figure 38d). Three libraries prepared from single blastocysts and one library prepared from dilute gDNA were submitted for pilot sequencing to assess the quality of the libraries.

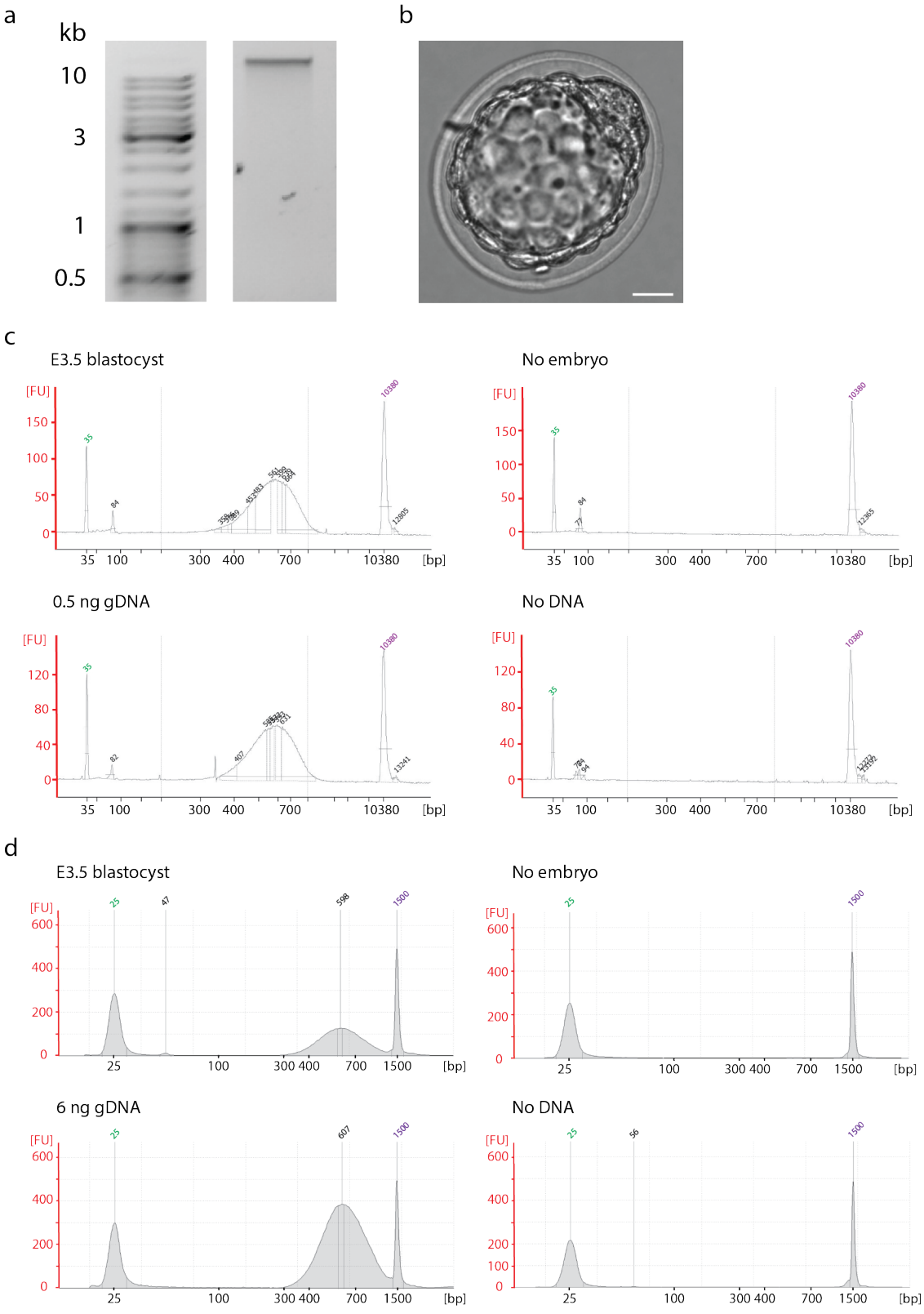

a
kb
10
3
1
0.5
b
c
E3.5 blastocyst
No embryo
0.5 ng gDNA
No DNA
[FU]
[bp]
d
E3.5 blastocyst
No embryo
6 ng gDNA
No DNA
[FU]
[bp]

Figure 38: Low-input BS-seq library preparation optimisation. Preparation of low-input BS-seq libraries. **a**: Purified genomic DNA used for optimisation libraries. **b**: Representative image of mouse E3.5 embryo used for optimisation libraries. Scale bar represents 20 μm. **c-d**: Bioanalyzer or TapeStation profiles of low-input BS-seq libraries from optimisation trials. Y axes show library concentration (fluorescence units; FU). X axes show library size distribution (base pairs; bp). Optimisation libraries were prepared from individual mouse E3.5 blastocyst, or from low amounts of purified gDNA as a positive control. 'No embryo' controls (picking buffer only) and 'no DNA' controls (library preparation buffers with only) were processed in parallel to control for contamination from exogenous DNA. **c**: trial libraries prepared using iPCRTag adapters. Size markers are shown in green (35 bp) and purple (10, 380 bp). **d**: trial libraries prepared using NEBNext adapters. Size markers are shown in green (25 bp) and purple (1500 bp).

100 bp paired-end sequencing yielded between 11 and 36 million reads per library (Table 9). Successful bisulfite conversion was confirmed by the cytosine depletion apparent in read one, and the reciprocal guanine depletion in read two (Figure 39a). In contrast to the low-input RRBS library presented in the previous section, the BS-seq pilot libraries had less than 10% adapter content (Figure 39b). Libraries were hard-trimmed to remove the biased sequence resulting from the 6N oligonucleotide used in first- and second-strand synthesis, as well as poor quality bases, and adapter read-through. Greater than 90% of bases were retained post-trimming (Table 9).

The trimmed reads were aligned to the mouse reference genome (mm10) using Bismark, resulting in mapping rates of 54-63% (Table 9). In RRBS libraries, multiple reads map at identical loci as a result of the restriction digestion used in library preparation. It is therefore impossible to determine if such reads are the result of PCR duplication or the amplification of genuine unique molecules, and therefore library deduplication is not appropriate. In contrast, BS-seq libraries derive from random fragmentation of the genome, and therefore reads mapping at the same locus are likely to derive from PCR duplication. Between 14-37% of reads were duplicate in the pilot BS-seq libraries, which were therefore deduplicated (Table 9).

Following mapping, between 6 and 31% of CpG sites were covered at least once. The library derived from adult gDNA had a higher coverage than the libraries derived from mouse embryos (6-10%, Figure 39c). The embryo libraries had a higher coverage than the pilot RRBS library, and the decay in coverage at higher thresholds implied that these

libraries were not sequenced to saturation, and that further CpGs could be captured by deeper sequencing (Figure 39c). Therefore, in comparison to low-input RRBS, BS-seq can capture a higher fraction of genomic CpGs, which when pooled *in silico* over replicate samples can provide a broader picture of the genomic methylation level.

The bulk CpG methylation level of the three blastocysts was between 16-20%, while the adult library had an overall CpG methylation level of 63% (Table 9). CHH methylation levels were below 3% in all libraries, implying that the bisulfite non-conversion rate was below 3% (Table 9).

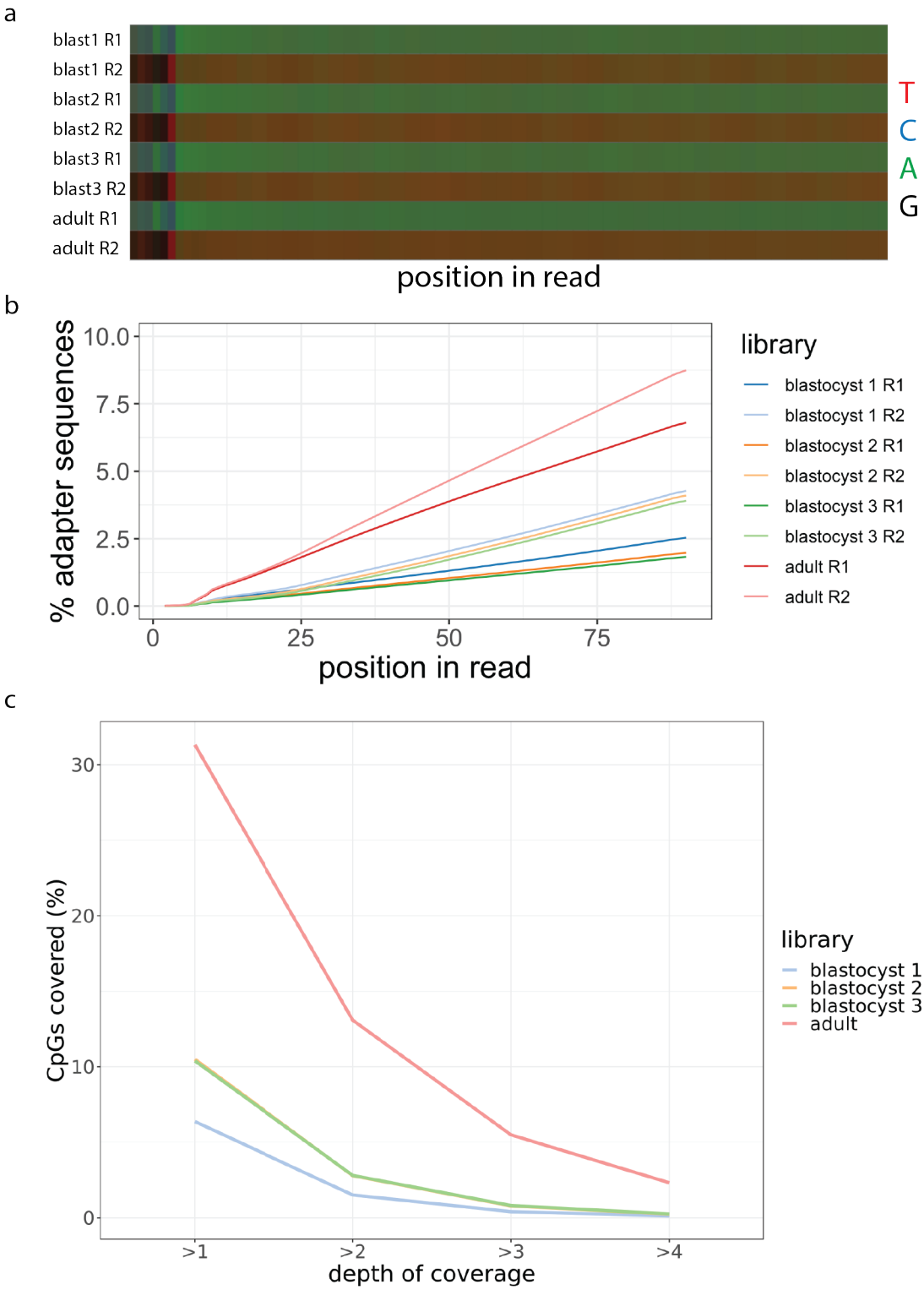


Figure 39: Post-sequencing quality control of pilot low-input BS-seq libraries. Post-sequencing assessment of the quality of the low-input Bs-seq libraries prepared from single mouse E3.5 blastocysts or low-concentration purified adult gDNA. **a**: Heatmaps depicting base composition for each read of pilot libraries. Each tile represents one base. (R1: read one; R2: read two). **b**: Cumulative percentage of adapter content as a function of position in the read. **c**: Percentage of CpG sites covered at different read depths.

Sample name	Number of reads (millions)	Mapping efficiency (%)	Bases retained post-trimming (%)	Duplicated sequences (%)	CpG methylation (%)	CHH methylation (%)
Blastocyst 1	11.6	63.95	96.35	36.92	19.85	2.45
Blastocyst 2	16.1	62.7	96.45	32.67	20	2
Blastocyst 3	17.5	63.1	96.6	37.64	16.2	1.9
Adult	35.1	54.9	93.85	14.68	63.1	2.85

Table 9: Low-input BS-seq library statistics.

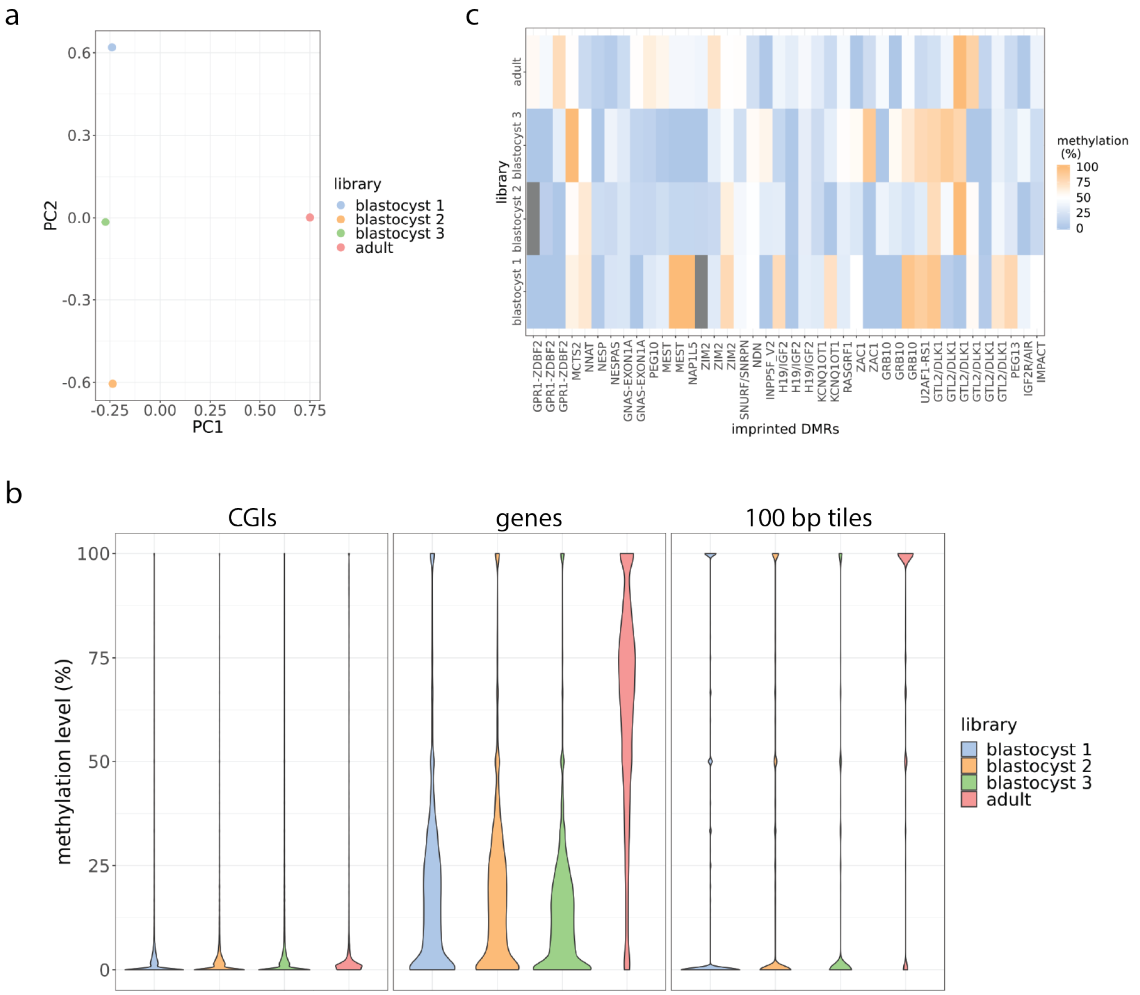


Figure 40: Global methylation patterns of pilot low-input BS-seq libraries.
a: Principle components analysis (PCA) showing the variation between pilot libraries for the top two principle components. The PCA was performed on methylation calls at individual CpG sites using the function 'PCASamples' in the R package "MethylKit'. n_{CpG} = 24405425. **b**: Violin plots representing the global methylation distribution of different genomic features. CGIs; UCSC defined CpG islands. n_{CGI} = 15877. Gene bodies: Ensembl defined genes from TSS to transcription end site. n_{gene} = 34873. 100 bp tiles; non-overlapping 100 bp tiles. n_{tile} = 9791637. **c**: Heatmap of methylation levels at 40 imprinting-associated differentially methylated regions (DMRs). DMRs; WAMIDEX defined regions converted to mm10 coordinates using the UCSC genome browser liftOver tool. DMR names are repeated in cases where multiple DMRs are associated with the same imprinted locus. Grey tiles indicate missing data.

A principle components analysis (PCA) of the four pilot libraries revealed that the three embryo libraries grouped together and separated from the adult library along principle component 1 (Figure 40a). Upon examining the methylation distribution of different genomic features, the expected hypomethylation of embryo libraries was apparent (Figure 40b). At CpG islands, both embryos and adult showed low methylation levels, whereas

at gene bodies, or 100 bp tiles covering the genome, the embryo libraries were hypomethylated relative to the adult. At imprinted DMRs, where the average of parental alleles is expected to result in an intermediate methylation level, embryo samples displayed varied patterns of methylation that were not always consistent between different embryos (Figure 40c). These results show that low-input BS-seq libraries can resolve global methylation levels, both in an unbiased approach like PCA, and in examining the methylation of particular genomic features. However, low-input BS-seq libraries may not capture the complete methylation profile of individual loci, perhaps due to allelic drop out or incomplete coverage, as can be expected in low-input sequencing approaches.

Low-input BS-seq libraries accurately recapitulated reported global methylation patterns of mouse embryos, and offered an improved coverage compared to low-input RRBS, as well as the ability to differentiate and remove probable PCR duplicates. I therefore proceeded to apply the low-input BS-seq technique to a timeline of opossum early development.

5.2 Generation of single-embryo BS-seq dataset covering a timeline of opossum early development

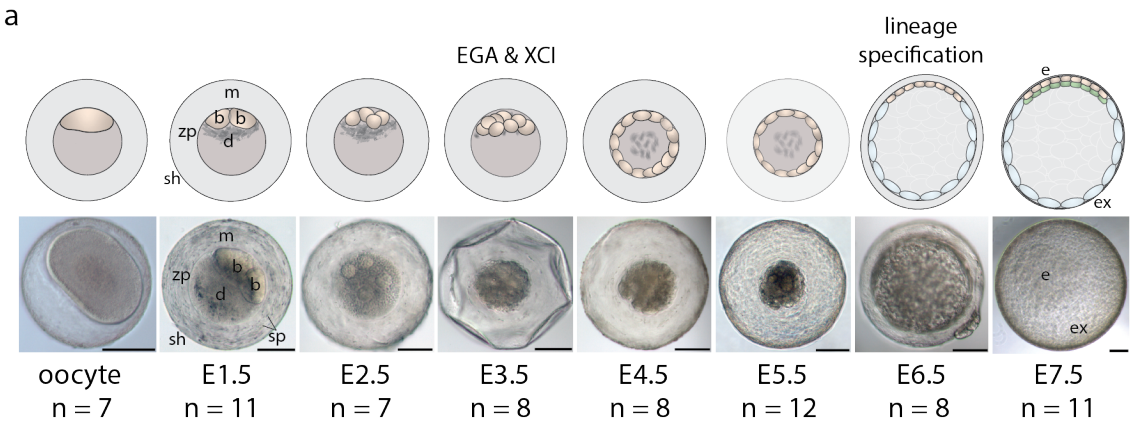
To study the role of DNA methylation in early opossum development, I profiled the methylation of oocytes, sperm, and a developmental timecourse from embryonic day 1.5 - 7.5 (Figure 41a). This timecourse encompasses developmental events from shortly after fertilisation (E1.5), through cleavage stages (E1.5-5.5), to the formation of the blastocyst (E6.5-7.5). This timecourse also captures the window during which EGA and XCI occur, defined by RNA-seq and RNA FISH experiments as beginning at E3.5 in the opossum (Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished). At E7.5, the embryonic and extra-embryonic lineages are morphologically distinguishable. RNA-sequencing experiments have shown that differences in cell lineage can be observed from E6.5 at the transcriptional level.

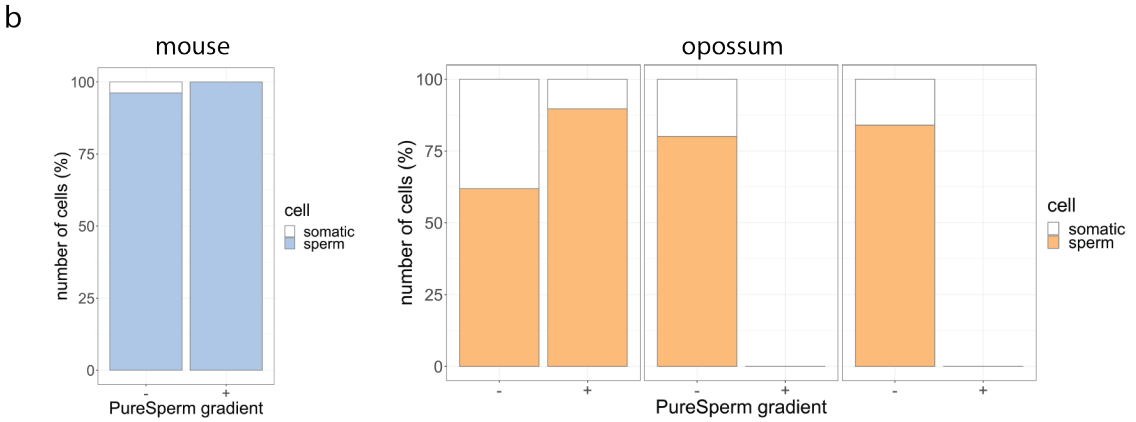
For the collection of oocytes, a female opossum was mated to a vasectomised male and oocytes were recovered from the uterus at E1.5 (Figure 41a). Variation in exact developmental staging is often observed between different animals when performing timed collections of opossum embryos (Mate et al., 1994). For example, E1.5 collections can yield zygotes or 2-cell stage embryos; in these samples most E1.5 embryos had reached cleavage (Figure 41a). For samples from E1.5-5.5, each timepoint represents the embryos recovered from the uterus of a single female; i.e. one litter. For E6.5 and E7.5, the samples represent embryos from two litters per timepoint. Opossum oocytes and embryos are surrounded by a mucoid layer and shell coat which are deposited during passage through the oviduct after ovulation, and in which sperm and maternal cells become trapped (Figure 41a). These outer layers were removed from all samples by protease digestion and manual dissection, followed by extensive washing, to ensure the complete removal of contaminating non-embryonic cells. Individual oocytes and embryos were then processed for library preparation following the BS-seq method optimised in section 5.1.2.

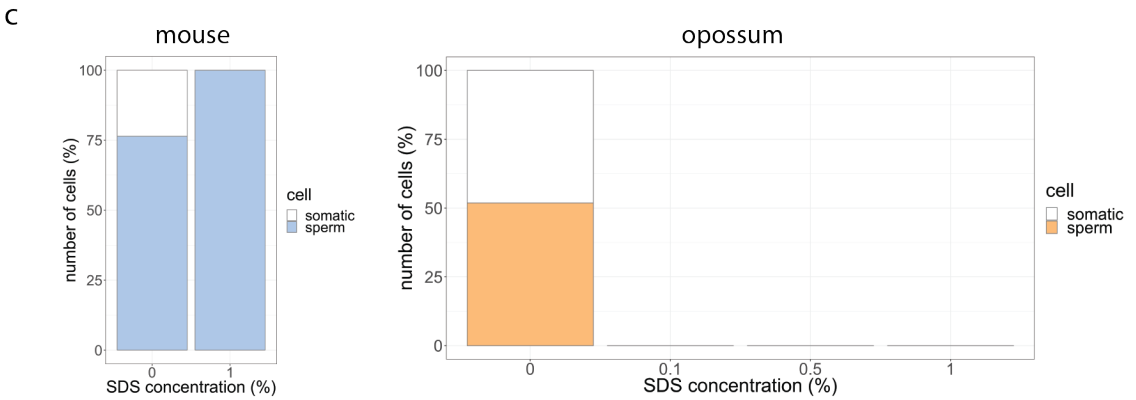
I tested several approaches for collecting opossum sperm, aiming to find a method which successfully excluded any contaminating non-sperm cells. I initially purified both mouse and opossum epididymal extract using the PureSperm gradient purification method (Figure 41b). The PureSperm method effectively removed somatic cells from the mouse sample (100% sperm cells; Figure 41b). When opossum samples were applied to the PureSperm gradient, results were varied; on occasions where cells were recovered post-gradient, the percentage of sperm in the sample had risen relative to before purification, though not to 100% as seen for mice (Figure 41b). However, it was frequently the case that no opossum cells were recovered from the gradient. Due to the variability in yield and the retention of somatic cells in the post-gradient sample, I concluded the PureSperm gradient was not effective for purification of opossum sperm.

I next tested an alternative technique based on the differential susceptibility of sperm and somatic cells to lysis in detergent buffer (Déborah Bourc'his, personal communication; Figure 41c). Incubation of mouse epididymal extract in buffer containing 1% SDS effectively lysed somatic cells while leaving sperm intact (Figure 41c). However, when

opossum epididymal extracts were incubated in 1% SDS buffer, both sperm and somatic cells were lysed (Figure 41c). I tested the effect of lowering the SDS concentration and found that opossum sperm were lysed at SDS concentrations as low as 0.1% (Figure 41c). Finally, I optimised an approach to manual picking of individual sperm from dilute suspensions of epididymal extract using a Stripper micropipette, followed by serial washes through fresh buffer (Figure 41d). Using this technique, I successfully collected samples of single sperm, and of ~ 30 pooled sperm cells. These sperm were then processed following the BS-seq method used above, with modifications including harsher cell lysis conditions and lengthened bisulfite conversion time, as recommended for preparation of sperm libraries (Stephen Clark, Reik laboratory, personal communication).


a
EGA & XCI
lineage specification
m
b
b
zp
d
sh
e
ex
sp
oocyte
n = 7
E1.5
n = 11
E2.5
n = 7
E3.5
n = 8
E4.5
n = 8
E5.5
n = 12
E6.5
n = 8
E7.5
n = 11


b
mouse
opossum
number of cells (%)
100
75
50
25
0
cell
somatic
sperm
-
+
PureSperm gradient


c
mouse
opossum
number of cells (%)
100
75
50
25
0
cell
somatic
sperm
0
1
0.1
0.5
SDS concentration (%)

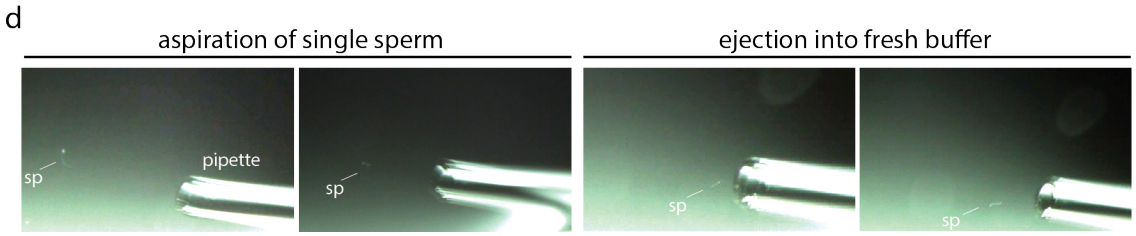

d
aspiration of single sperm
ejection into fresh buffer
pipette
sp

Figure 41: Collection of a timeline of opossum early development for low-input BS-seq. Opossum embryo and gamete samples collected for low-input BS-seq. **a**: Schematics (upper panels) and microscope images (lower panels) representing opossum oocytes and embryos from E1.5–7.5. Schematics are side views with the animal pole to the top. Scale bars represent 100 μm. Indicated on the E1.5 embryo image are the blastomeres (b), deutoplasm (d), zona pellucida (zp), mucoid coat (m), sperm trapped in the mucoid coat (sp), and the shell coat (sh). Indicated on the E7.5 embryo image are the embryonic (e) and extra-embryonic (ex) regions. Embryonic genome activation (EGA), X chromosome inactivation (XCI), and lineage specification are noted above the images of the timepoints at which they occur. **b** – **c**: Purity of sperm preparations resulting from different methods of removing somatic cells from samples of mouse (left) and opossum (right) epididymal sperm. For percentage of sperm quantifications, a minimum of 100 cells were counted. **b**: PureSperm gradient purification. Three independent replicates are shown for opossum. **c**: Differential lysis in detergent buffer. **d**: Video stills showing aspiration and dispensing of single opossum sperm using a Stripper micropipette.

100 bp paired-sequencing of libraries yielded between 85 – 165 million reads (Table 10). Alignment to the opossum reference genome using Bismark resulted in mapping rates of between 50-60% for libraries from E1.5-7.5 samples (Table 10). Oocyte and sperm libraries had mapping rates between 25-40% (Table 10). A large variation in mapping rate was observed for E4.5 embryos (Table 10). This variation was found to result from one library with a mapping efficiency of only 8%. This library was therefore excluded from subsequent analyses. I also observed variability in mapping rate in sperm libraries, which was attributable to higher mapping efficiencies in the libraries prepared from pooled sperm than the single-cell libraries. No individual sperm library had a notably poor mapping rate, and therefore all were retained for subsequent analyses. Duplication rates between 80-95% were observed in libraries derived from timepoints with few cells per sample (Table 10). Duplication rates were lower in libraries derived from E5.5 - 7.5 embryos, which was expected because the higher amount of input material available from larger embryos can increase library complexity (Table 10). Duplicated sequences most likely derive from PCR amplification and were therefore removed using the deduplication function in the Bismark package.

Between 40-50 % of reads contained adapter sequences (Table 10). Adapter content was not correlated with sample timepoint, and following trimming for quality and adapter content, greater than 87% of bases were retained across all samples (Table 10). Non-CpG (CHH) methylation levels were approximately 2-3% in all timepoints except oocytes,

where CHH methylation was 4.5%. This implied that in most libraries, the bisulfite non-conversion rate was no higher than 3%. The oocytes of eutherian mammals are known to harbour elevated levels of non-CpG methylation; it is possible the same effect occurs in opossum and accounts for the elevated CHH methylation observed here (Tomizawa et al., 2011).

Sample timepoint	Number of libraries	Millions of reads (sd)	% Mapping rate (sd)	% Duplicated sequences (sd)	% Reads with adapter content (sd)	% Bases retained after trimming (sd)	% CpG methylation (sd)	% CHH methylation (sd)
1.5 dpc	11	129.28 (29.54)	50.82 (4.40)	88 (4.29)	47.54 (3.04)	92.18 (4.17)	63.64 (2.01)	2.09 (0.30)
2.5 dpc	7	96.33 (13.98)	53.14 (3.67)	88 (2.45)	44.85 (0.37)	96.57 (0.53)	66.43 (1.62)	2.00 (0.00)
3.5 dpc	8	165.36 (24.56)	55.38 (2.83)	92.50 (2.93)	43.25 (0.71)	98 (0)	64.50 (3.07)	1.88 (0.35)
4.5 dpc	8	116.64 (8.60)	48.63 (17.55)	91.25 (3.11)	42.37 (1.76)	97.38 (0.52)	65.75 (6.04)	3.25 (2.82)
5.5 dpc	12	130.61 (14.65)	55.25 (1.96)	61.75 (9.58)	42.17 (0.39)	97 (0)	49.83 (2.08)	2.25 (0.45)
6.5 dpc	8	86.09 (12.29)	58.50 (1.41)	20.50 (4.21)	51.75 (4.17)	91.13 (3.44)	43.13 (1.89)	1.88 (0.64)
7.5 dpc	11	95.42 (16.64)	59.73 (1.95)	22.45 (6.39)	55.27 (8.96)	87.28 (7.73)	47.45 (2.98)	2.18 (0.60)
Oocyte	7	71.61 (18.36)	25.25 (4.98)	84.50 (5.68)	41.63 (2.29)	96.75 (0.46)	71.13 (3.04)	4.50 (1.20)
Sperm	3 single sperm 4 pools (~30 sperm)	117.39 (29.88)	39.71 (24.54)	95.71 (3.09)	42.71 (2.29)	96.71 (0.49)	71.57 (2.30)	1.43 (0.53)

Table 10: Average library statistics per timepoint for opossum low-input BS-seq libraries.

To facilitate analysis of DNA methylation on the X chromosome of opossum embryos, it was necessary to determine the sex of each embryo sample. It was not possible to determine the sex of the embryos *a priori*, as the timepoints profiled were prior to development of physical sex characteristics. Sex can be assigned to samples post-sequencing by counting the number of reads mapping to the X and Y chromosomes (Figure 42a). The publicly available *M. domestica* reference genome assembly does not contain Y chromosome sequence (Mikkelsen et al., 2007b), so a pseudoY chromosome was constructed using known Y chromosome CDS (Bellott et al., 2014, Cortez et al., 2014). Opossum embryos segregated into two groups based on X- and Y- mapped reads; 37 embryos had mapped Y reads and approximately half the number of X reads as other embryos, and were designated as male (57% of embryos). 28 embryos had negligible Y reads and approximately twice the number of X reads, and were designated as female (43% of embryos, Figure 42a). Oocytes do not contain a Y chromosome and had negligible Y-mapped reads (Figure 42a). The pooled sperm libraries could contain both X- and Y-carrying sperm, and in accordance with this I observed sperm libraries with reads mapping to both X and Y chromosomes (Figure 42a). This analysis also underscores the successful removal of any contaminating cells from the mucoid layer of embryo samples, as retention of sperm or maternal cells from this structure would have resulted in embryos with high mapping rates to both X and Y (Figure 42a).

PCA revealed that methylation variation among samples primarily occurred according to timepoint (Figure 42b). The most obvious grouping was the separation of sperm samples from oocytes and embryos along PC1 (Figure 42c). Embryo and oocyte samples grouped together, with some separation by timepoint along PC2, for example E6.5 and E7.5 samples (Figure 42b).

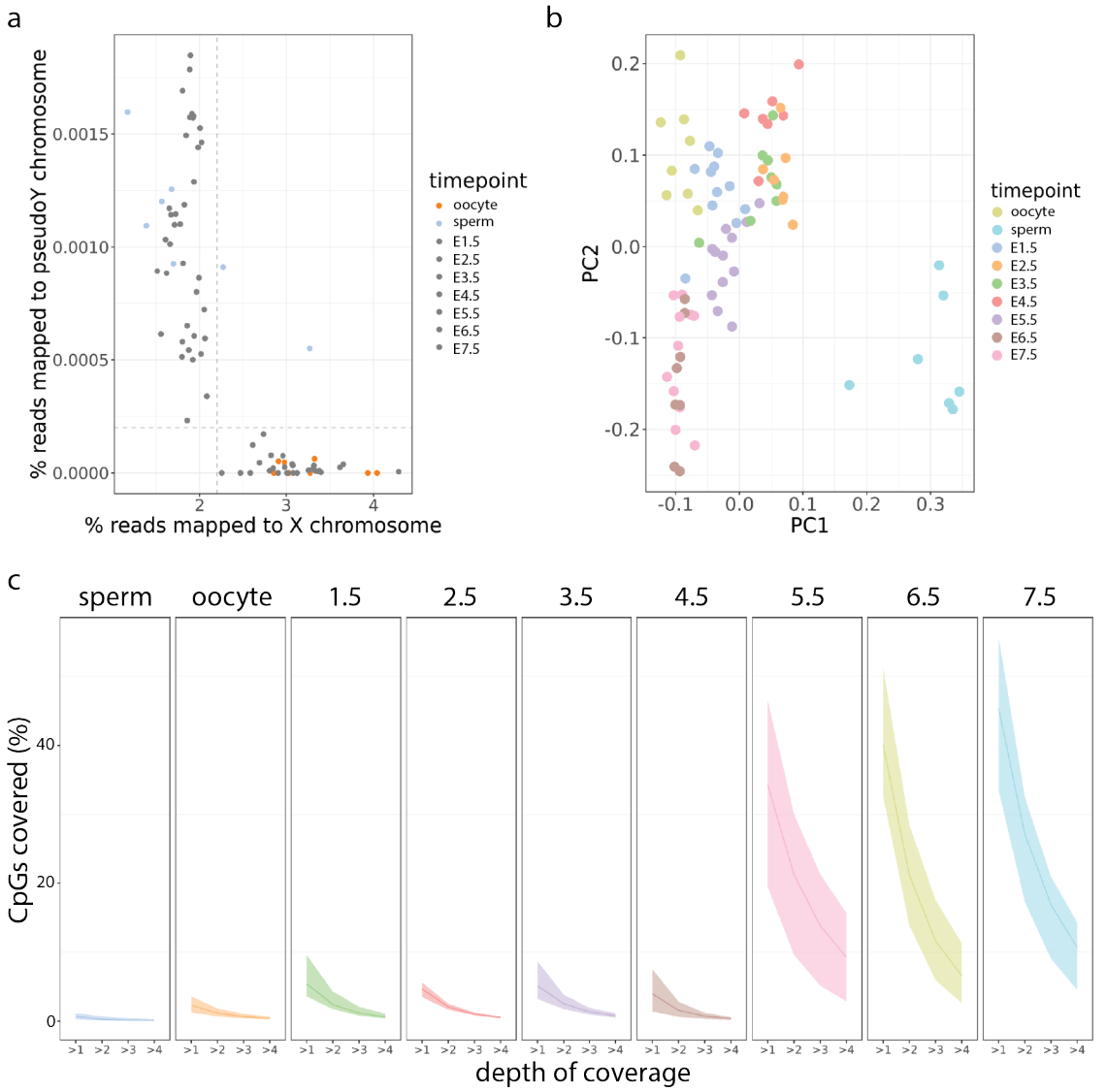


Figure 42: Post-sequencing quality control of opossum BS-seq libraries. a: Percentage of reads mapped to the X and pseudoY chromosomes in each embryo library. Grey dashed lines indicate thresholds chosen for assignment of sex to samples. **b**: Principle components analysis (PCA) depicting the variation in methylation between each library for the top two principle components. The PCA was performed on methylation calls at all CpG sites covered per library using the function 'PCASamples' in the R package "MethylKit'. **c**: Percentage of CpG sites covered at different read depths. Line indicates the mean per timepoint, and ribbons indicate the maximum and minimum values.

I calculated the percentage of total genomic CpGs that were covered in each BS-seq library (Figure 42c). Coverage increased with developmental stage, with gamete libraries capturing between 1-2% of CpG sites, stages 1.5-4.5 capturing 4-5%, and stages 5.5-7.5 capturing between 34-45% of CpGs (Figure 42c). The differences in coverage between timepoints can be attributed to increased library complexity in the samples with greater input material, similar to my observations above regarding duplication rates. To further analyse DNA methylation patterns in these data, I pooled the data by timepoint, creating an *in silico* sample for each stage. These datasets were filtered to include only CpG sites present in all timepoints with a minimum coverage of 3 reads. The resulting dataset represented 26,077 CpG sites.

In order to understand the distribution of these 26,077 CpGs throughout the genome, I examined the intersection of the captured CpGs with different genomic features (Figure 43). Approximately 25% of CpG sites overlapped CpG islands, with this distribution split nearly evenly between intergenic, intragenic, and promoter CGIs (Figure 43a). Nearly 20% of CpGs fell inside either genes or their promoters (Figure 43b). 52% of CpGs were located on autosomes, with nearly 1% located on the X chromosome (Figure 43c). The remaining 42% of sites were located on chromosome 'Un', the pseudochromosome housing unassembled sequences (Figure 43c). 28% of CpGs did not overlap a RepeatMasker annotated repeat element (Figure 43d). Of the remaining CpGs, the majority overlapped LINE or LTR elements (Figure 43d).

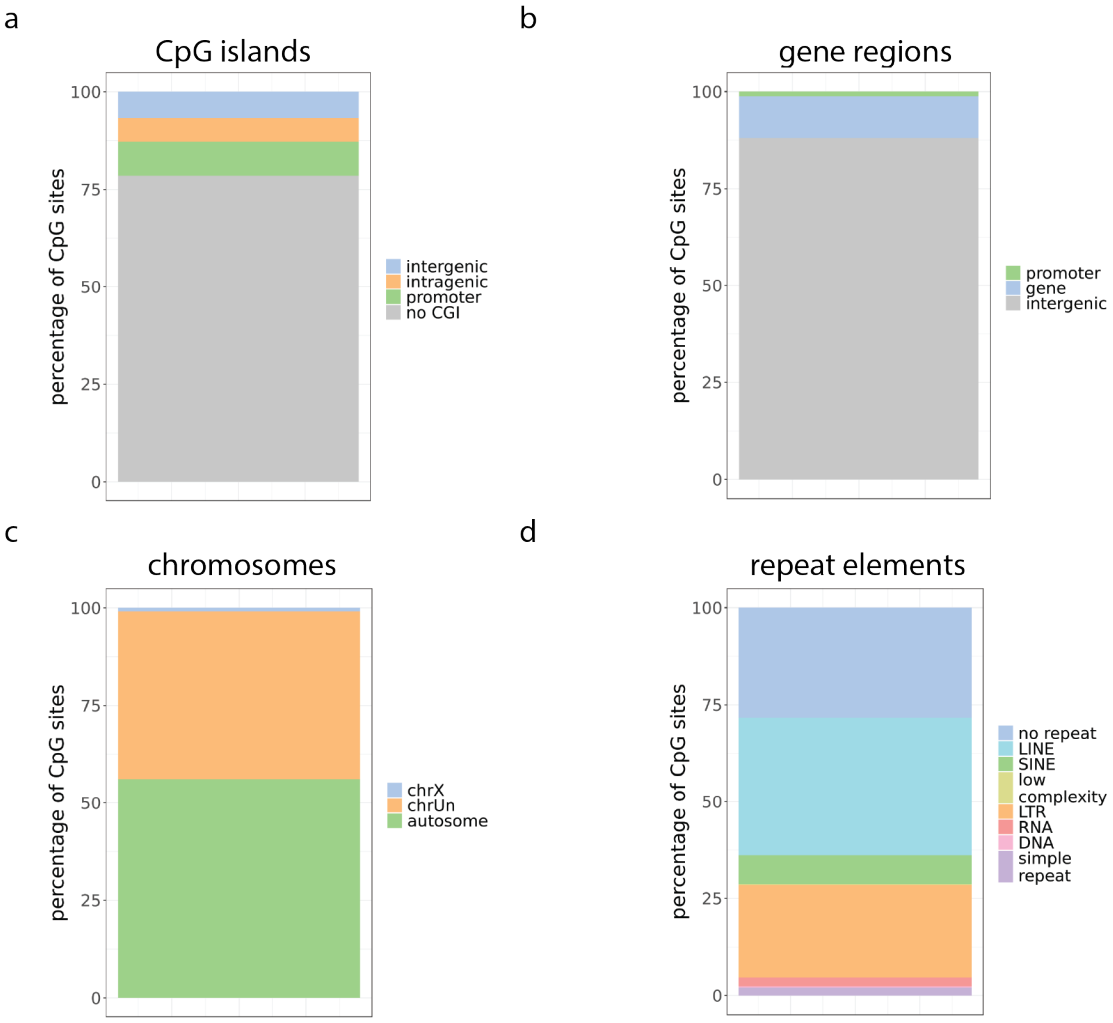


Figure 43: Genomic distribution of CpG sites captured by low-input BS-seq libraries. Distribution across different genomic features of the 26,077 CpG sites covered in the pooled opossum embryo BS-seq dataset. **a**: Distribution of CpG sites in CpG islands. **b**: Distribution of CpG sites in genes and promoters. **c**: Distribution of CpG sites across chromosomes. **d**: Distribution of CpG sites across different types of repeat element. Note that the different genomic features are not mutually exclusive categories.

5.3 Global methylation dynamics during opossum development

To examine methylation dynamics during opossum development, I compared the global distribution of methylation at the 26,077 common CpGs between gametes and embryos from E1.5-7.5 (Figure 44a). Methylation levels at the same CpG sites in adult tissues were examined for comparison (Figure 44a; BS-seq data generated in Chapter 4). This analysis revealed that at all timepoints profiled, methylation was bimodally distributed with a

greater number of sites showing high methylation than low methylation. This bimodal pattern was reminiscent of the global pattern seen in adult opossum tissues (Figure 44a).

Although all opossum early developmental timepoints showed a bimodal distribution, some differences in methylation were apparent between stages. Sperm had a median methylation level of 90%, higher than that observed for oocytes, where the median methylation was 65% (Figure 44a). The median methylation level of E1.5 embryos was 74%, a level approximately intermediate between sperm and oocyte. For the subsequent three days of development, methylation levels remained largely similar to E1.5, with median levels between 70-76% (Figure 44a). From E5.5 -7.5, methylation levels shifted downwards, with fewer sites displaying 100% methylation, and median levels between 53-62% (Figure 44a).

A heatmap clustering individual CpGs according to methylation level revealed the different dynamics of methylation during development in more detail (Figure 44b). A large proportion of CpG sites remain hypermethylated at every stage of development (Figure 44b). Another large subset of CpGs remain hypomethylated during the entire timecourse (Figure 44b). This largely hypermethylated genome contrasts to the global hypomethylation observed in pre-implantation mouse embryos (Figure 44c). In addition to the regions where methylation remained constant across development, smaller clusters were also apparent, at which methylation patterns were more dynamic. Across most clusters, changing methylation had resolved to an adult-like pattern by the E6.5-7.5 embryonic stages (Figure 44b).

Of the clusters with dynamic methylation, two related to CpG sites at which the methylation state differed between the gametes. One such group of sites were highly methylated in sperm and lowly methylated in oocytes, passed through a phase of intermediate methylation in early embryos, and resolved to low methylation in later embryonic stages and adulthood (Figure 44b). A smaller cluster of CpGs displayed the opposite trend; these were lowly methylated in sperm and highly methylated in oocytes, passed through an intermediate level in early embryos and resolved to low methylation in adults (Figure 44b). Interestingly, I did not observe clusters at which differing methylation states in the gametes resolved to high methylation in adults. However,

several clusters were observed at which methylation was initially similar between the gametes, passed through a phase of intermediate methylation, and resolved to either hypo- or hypermethylation in adults (Figure 44b).

In order to directly compare sites between timepoints, these analyses used the stringent set of 26,077 CpGs common to all samples. However, the trends seen in these data were also observed when considering the average bulk CpG methylation levels (Table 10), calculated without filtering for common sites between the samples. The similarity between the stringent and bulk approaches suggests that the conclusions drawn from stringent sites are representative of the remainder of the genome. These results suggest that in opossums, unlike eutherian mammals, there is not an immediate global demethylation following fertilisation, nor a sustained period of global hypomethylation during early development. Instead, these results point to a largely stable embryonic methylation landscape punctuated by regions of local reprogramming to an adult-like state, similar to the embryonic methylation landscapes described for the zebrafish (Potok et al., 2013, Jiang et al., 2013).

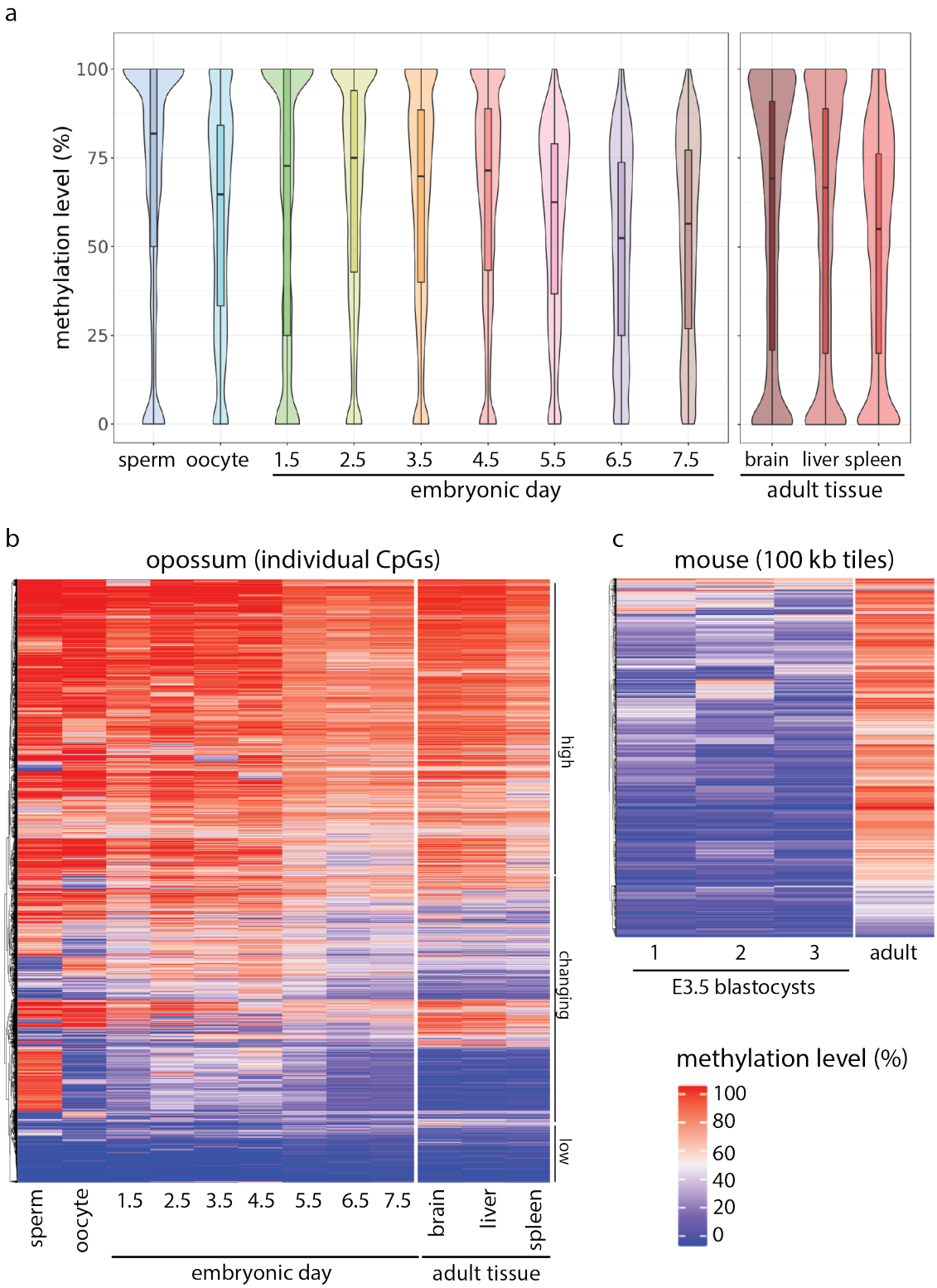

a
methylation level (%)
100
75
50
25
0
sperm
oocyte
1.5
2.5
3.5
4.5
5.5
6.5
7.5
embryonic day
brain
liver
spleen
adult tissue
b
opossum (individual CpGs)
high
changing
low
sperm
oocyte
1.5
2.5
3.5
4.5
5.5
6.5
7.5
embryonic day
brain
liver
spleen
adult tissue
c
mouse (100 kb tiles)
1
2
3
adult
E3.5 blastocysts
methylation level (%)
100
80
60
40
20
0

Figure 44: Global methylation dynamics during opossum development.
Distribution of methylation during a timecourse of early opossum development. Each developmental stage represents the *in silico* pool of all libraries from that timepoint. Methylation level was calculated as the percentage of methylated CpGs/ covered CpGs per site. Data represent 26,077 CpG sites that were covered at least 3x in every embryo timepoint. 26059 of these CpG sites were covered in the adult opossum BS-seq data generated in Chapter 4, and are presented here as a comparison to the embryo data. Male and female samples were combined per tissue. **a**: Violin plots representing the overall methylation distribution. Nested boxplots indicate the median methylation and 1st and 3rd quartiles. **b**: Heatmap representing the methylation level of individual CpG sites. **c**: Heatmap representing the methylation level of 100 kb genomic tiles in mouse E3.5 blastocysts and adult tissue (data replotted from section 5.1.2). Heatmaps were constructed using the R package 'complexHeatmap' by manually ordering according to sample timepoint on the x-axis, and permitting clustering by Euclidian distances on the y axis, using the hclust method "complete".

5.4 Methylation dynamics at genomic features during opossum development

To further explore those regions of the genome where methylation levels appeared to be changing during early development, I examined sites displaying significant differential methylation between consecutive developmental stages. By comparing each consecutive pair of samples, differential methylation was calculated for all CpG sites covered by at least 3 reads in both samples being compared, rather than only the 26,077 stringent sites, providing a greater number of sites for analysis. I calculated the percentage of CpG sites with methylation changes between consecutive timepoints (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test; Figure 45a). The comparison between sperm and oocyte showed the greatest number of changing sites, followed by the sperm – E1.5 transition (Figure 45a). All other consecutive comparisons had very few significantly changing sites (Figure 45a). These results indicate that the majority of CpG sites were unchanging between consecutive timepoints, supporting the qualitative interpretation made above by examining global methylation distributions (Figure 44).

To further understand the regions that differed between sperm and oocyte, I examined their genomic location. The majority of sperm:oocyte differing sites were located on chrUn (Figure 45b). Sites that were highly-methylated in oocytes and lowly-methylated in sperm were located in all types of genomic feature examined, including CGIs, gene bodies, promoters, repeat elements, and intergenic regions (Figure 45c). In contrast, sites that were lowly methylated in oocytes and highly methylated in sperm were almost exclusively located in CGIs and intergenic regions (Figure 45c).

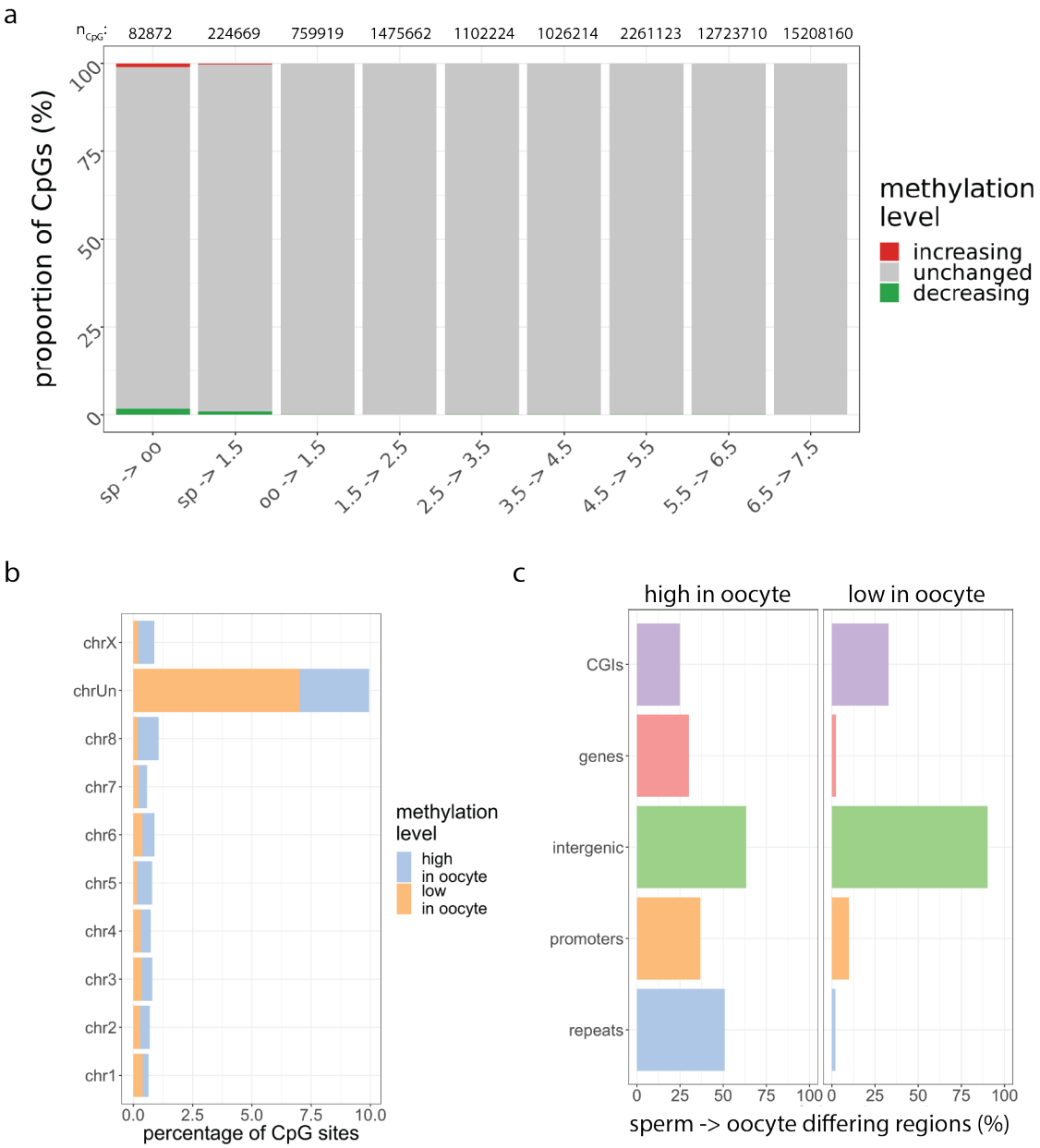


Figure 45: Differentially methylated regions in early opossum development. a: Percentage of CpG sites found to be differentially methylated between timepoints (methylation difference ≥ 25%, q-value <0.01, Fisher's exact test). Red: increased methylation, green: decreased methylation, grey: methylation unchanged. The number of CpG sites included in each comparison is indicated above. sp: sperm; oo: oocyte, 1.5-7.5: embryonic days of development. **b**: Distribution of differentially methylated regions between sperm and oocyte across different chromosomes. **c:** Percentage of differentially methylated regions between sperm and oocyte found to overlap different genomic features. CpG islands: UCSC defined CpG islands. Promoters: Ensembl defined transcriptional start sites (TSS) – 2kb/+ 200 bp. Gene bodies: Ensembl defined TSS to transcriptional end site. Intergenic regions: all regions not contained within a gene body. Repeats: RepeatMasker annotated repetitive elements.

In order to explore the methylation dynamics of different genomic features across opossum early development, I overlapped the stringent set of 26,077 CpGs with different feature annotations and calculated the methylation level per feature. A general decrease in methylation level was evident in E5.5-7.5 samples in all genomic contexts. CpG islands displayed a bimodal methylation pattern at all timepoints analysed (Figure 46a), with most showing no change between timepoints. However, a small group of CGIs were differentially methylated in gametes and were gradually reprogrammed during development to reach an intermediate methylation level in later embryos and adults (asterisk, Figure 46a). Another group of CGIs passed through a period of intermediate methylation in embryos before resolving to a gamete-like state in adult tissues (asterisk, Figure 46a). Very similar trends were observed at promoters, which is expected due to the significant co-occurrence of CGIs with promoters (asterisk, Figure 46b).

Most gene bodies were highly methylated, and few showed large changes across development (Figure 46c). A smaller group of gene bodies was hypomethylated at all timepoints examined (Figure 46c). A further small proportion switched from hypomethylation during embryonic stages to hypermethylation in adult tissues (asterisk, Figure 46c). At intergenic regions and RepeatMasker repeat elements, most sites were hypermethylated consistently throughout development, though the general trend toward a slight hypomethylation at E5.5-7.5 was also observed at these regions (Figure 46d-e). A small subset of intergenic regions showed a dynamic methylation pattern wherein gametes showed different methylation profiles, early embryos harboured intermediate methylation, and methylation levels resolved to low levels in adult tissues (asterisk, Figure 46d). A small subset of repeat regions, including some LINEs, SINEs, and LTRs displayed a similar dynamic pattern (asterisk, Figure 46e).

In summary, this survey of global methylation patterns in gametes and early embryos suggested that no global hypomethylation phase exists in opossum early development. In embryonic stages E5.5 -7.5, the global methylation level was slightly lower than in earlier embryonic stages or adult tissues. Examination of methylation dynamics of different genomic regions reinforced the conclusion that most sites did not undergo a dramatic methylation change, while revealing a small subset of regions where local reprogramming

occurred. Those sites where methylation underwent changes during development had mostly resolved to an adult-like pattern by E7.5.

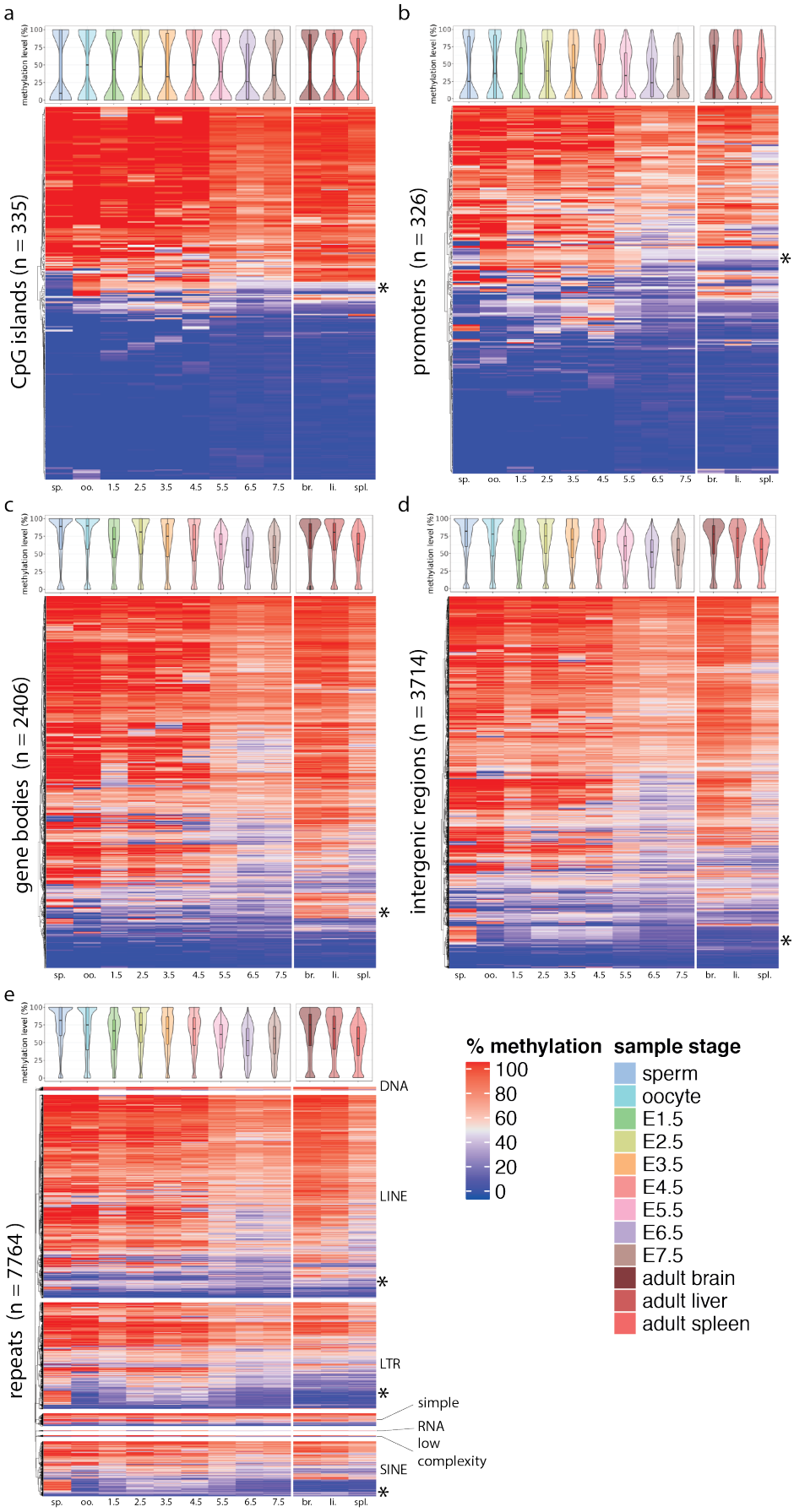

a
b
c
d
e
methylation level (%)
CpG islands (n = 335)
promoters (n = 326)
gene bodies (n = 2406)
intergenic regions (n = 3714)
repeats (n = 7764)
sp. oo. 1.5 2.5 3.5 4.5 5.5 6.5 7.5 br. li. spl.
DNA
LINE
LTR
simple
RNA
low
complexity
SINE
% methylation
100
80
60
40
20
0
sample stage
sperm
oocyte
E1.5
E2.5
E3.5
E4.5
E5.5
E6.5
E7.5
adult brain
adult liver
adult spleen

Figure 46: DNA methylation dynamics at genomic features across opossum development. Distribution of methylation at different genomic features during a timecourse of early opossum development. Each heatmap tile represents an individual genomic feature; violin plots above heatmaps represent the overall distribution per timepoint. Heatmap was constructed using the R package 'complexHeatmap' by manually ordering according to sample timepoint on the x-axis, and permitting clustering by Euclidian distances, using the hclust method "complete". Methylation level was calculated as the percentage of methylated CpGs/ covered CpGs within a region. Data represent 26,077 CpG sites that were covered at least 3x in every embryo timepoint. 26059 of these CpG sites were covered in the adult opossum BS-seq data generated in Chapter 4, and are presented here as a comparison to the embryo data. Methylation values for adult data are the average of three male and three female samples for each tissue. sp.: sperm, oo.: oocyte, 1.5-7.5: embryonic days of development, br.: brain, li.:liver, spl.: spleen. **a**: CpG islands (UCSC defined CpG islands). **b**: Promoters (Ensembl defined transcriptional start sites (TSS) – 2kb/+ 200 bp). **c**: Gene bodies (Ensembl defined TSS to transcriptional end site). **d**: Intergenic regions (all regions not contained within a gene body). **e**: RepeatMasker annotated repetitive elements.

5.5 Methylation levels in opossum pronuclear zygotes

The low-input BS-seq data presented above suggested that unlike eutherian mammals, the opossum genome does not undergo a rapid demethylation immediately following fertilisation. In order to examine the methylation of the two pronuclei separately at the zygote stage, I investigated bulk methylation levels by immunostaining with a 5-methylcytosine (5mC) antibody. This approach also served to validate the observations made by low-input BS-seq via an independent experimental technique.

Initially, I validated the 5mC antibody by staining adult testis and ovary sections. Gonad sections were co-stained for the germ cell marker DDX4, using an antibody previously confirmed to recognise eutherian and metatherian DDX4 (Hickford et al., 2011; Figure 47). As expected based on previous reports, DDX4 marked spermatogonia and spermatocytes within the seminiferous tubule, but was absent from elongating spermatids in both mouse and opossum testes (Fujiwara et al., 1994, Toyooka et al., 2000, Hickford et al., 2011; Figure 47a). DNA methylation staining could be observed within cell nuclei, for example marking spermatids in the lumen (white arrowheads) as well as the nuclei of spermatocytes (Figure 47a). Developing oocytes were marked by DDX4 in the ovaries of both mouse and opossum, as previously reported (Toyooka et al., 2000, Hickford et al., 2011; Figure 47b). DNA methylation could be faintly observed within the nuclei of the developing oocytes (white arrowheads; Figure 47b). DNA methylation staining was observed more strongly in the nuclei of surrounding somatic cells (Figure 47b).

Subsequently, the 5mC antibody was applied to opossum oocytes collected from the ovary at approximately 24 hpc, to establish conditions appropriate for whole-mount immunostaining (Figure 47c). Oocytes were co-stained with an antibody recognising H3K9me3 as a marker of chromatin (Figure 48a). As expected for DNA-associated epitopes, the H3K9me3 and 5mC signals colocalised on the condensed meiotic chromosomes (inset, Figure 47c), confirming the whole-mount immunostaining conditions were successful.

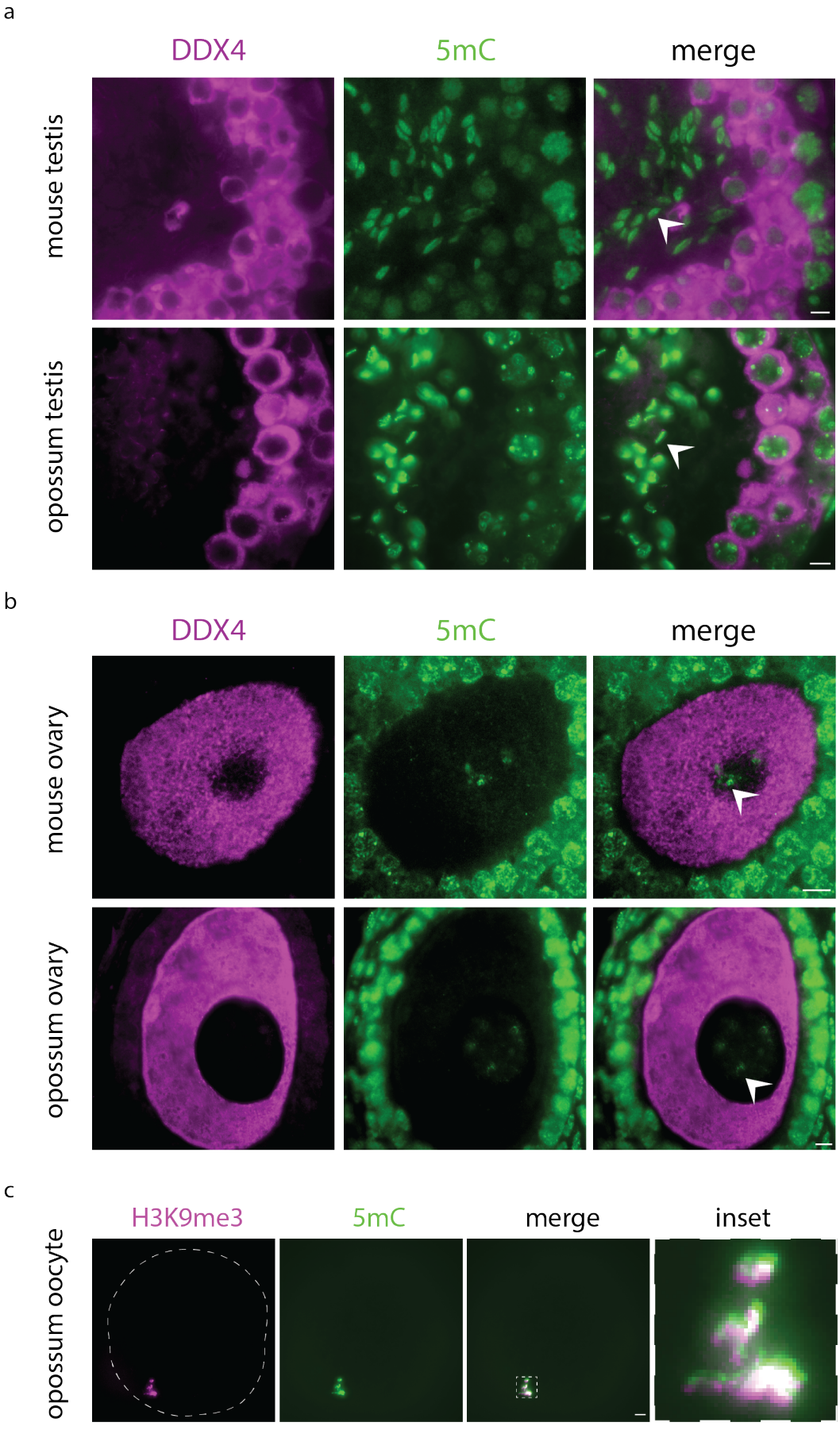

a
DDX4
5mC
merge
mouse testis
opossum testis
b
DDX4
5mC
merge
mouse ovary
opossum ovary
c
H3K9me3
5mC
merge
inset
opossum oocyte

Figure 47: Optimisation of 5mC immunostaining.
a-b: Mouse and opossum adult gonadal tissue sections stained for DDX4 (magenta) to mark germ cells and 5mC (green) marking DNA methylation within cell nuclei. **a:** Ovary; views show one developing oocyte and surrounding follicle. White arrowheads indicate DNA methylation within oocyte nuclei. **b:** Testis; views show part of one seminiferous tubule with spermatid nuclei in the lumen (white arrowheads). **c:** Whole-mount opossum oocyte stained for H3K9me3 (magenta) and 5mC (green). Representative image, a total of 3 oocytes were stained. Inset shows meiotic chromosomes. Scale bars represent 10 μm.

Subsequently, 26 and 29 hpc opossum zygotes were collected and stained by Sugako Ogushi (Turner laboratory), following the same procedure used for oocytes (Figure 48a). In addition, zygotes were counterstained with propidium iodide to visualise DNA. The 5mC and H3K9me3 signals were quantified for each pronucleus, and the signal difference between pronuclei calculated for each embryo (Figure 48b). In 26 hpc zygotes, both pronuclei were positive for H3K9me3, but the staining was asymmetric, with one notably H3K9me3-enriched pronucleus in each zygote (Figure 48). By 29 hpc, H3K9me3 staining was largely symmetrical between pronuclei (Figure 48). In contrast to H3K9me3, 5mC staining was equally strong in both pronuclei at 26 and 29 hpc (Figure 48).

In the mouse, demethylation of the paternal pronucleus can be observed by 5mC immunostaining after fertilisation (Santos et al., 2002). The pronuclei of mouse zygotes are also asymmetrically enriched for H3K9me3, with the maternal pronucleus marked by high H3K9me3 signal and the paternal pronucleus remaining devoid of H3K9me3 until after the two-cell stage (Santos et al., 2005). There is as yet no defined pronuclear staging scheme for opossum zygotes, and it is therefore unclear how the opossum zygotes shown here relate to substages of the mouse zygote. However, it is likely that the H3K9me3-enriched pronucleus at 26 hpc represents the presumptive maternal pronucleus. The similarity in 5mC signal between pronuclei at both 26 and 29 hpc implies that the paternal genome is not subject to rapid demethylation in the opossum zygote, supporting the findings from low-input BS-seq in section 5.3.

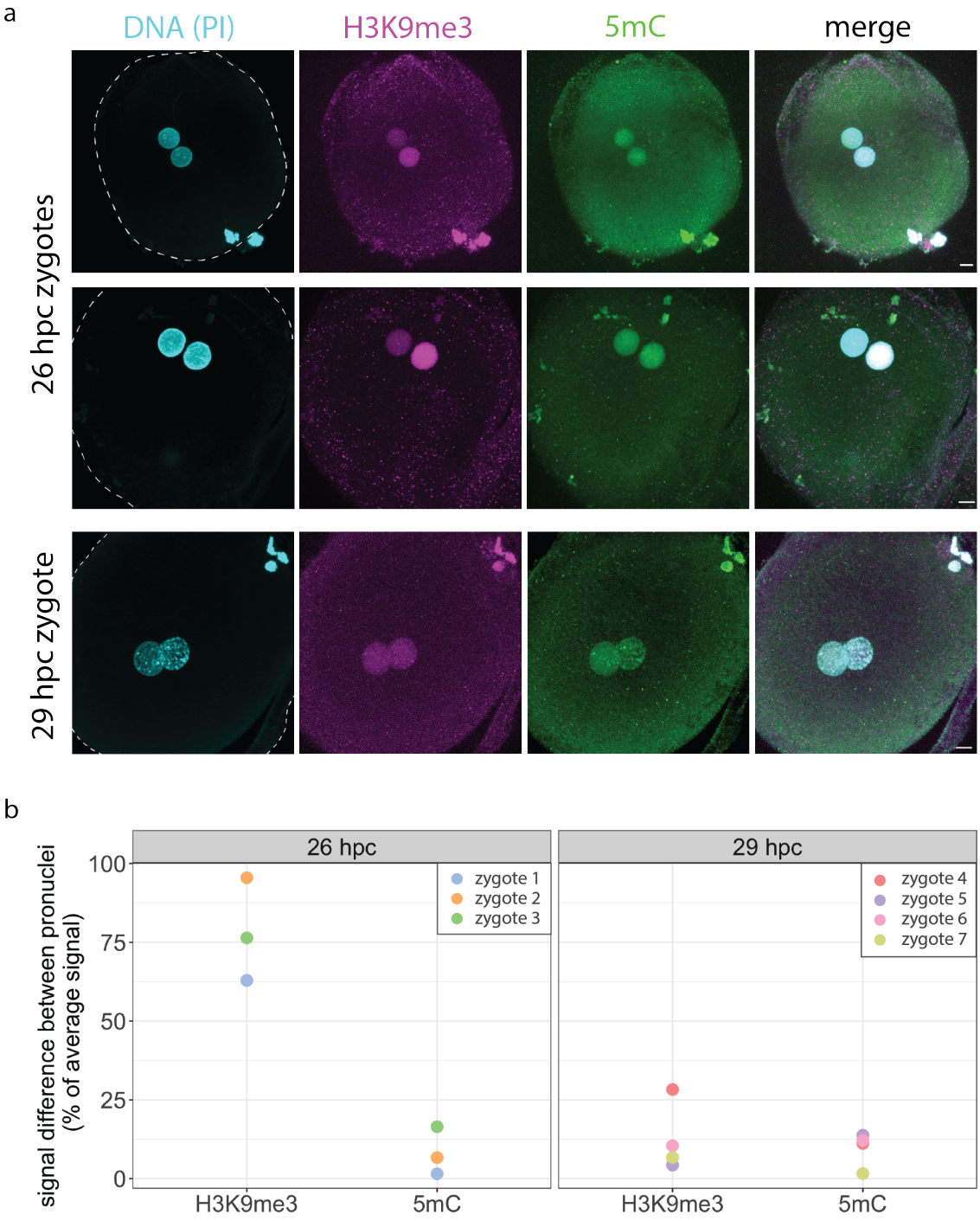


Figure 48: 5mC immunostaining of opossum oocytes and zygotes.
a: Opossum zygotes stained for H3K9me3 (magenta) and 5mC (green). Zygotes collected at 26 and 29 hours post-coitum (hpc). DNA was counterstained with Propidium iodide (PI; cyan). Two representative images are shown for 26 hpc, and one for 29 hpc. Total zygotes stained: $n_{26\ hpc} = 3$, $n_{29\ hpc} = 4$. Scale bars represents 10 µm. **b:** The difference in signal intensity between pronuclei (pn) for the 5mC and H3K9me3. Signal difference was calculated as $pn_1 - pn_2/(pn_1 + pn_2)/2$. Zygote staining and quantification of pronuclei were performed by Sugako Ogushi, Turner laboratory.

5.6 Expression of methylation enzymes during opossum development

Following the observation that there was not a drastic demethylation of the opossum genome in the early developmental stages, I was interested to examine the expression of opossum orthologues of enzymes involved in DNA methylation. To examine expression at embryonic stages, I made use of an opossum embryo single-cell RNA-seq (scRNA-seq) dataset generated by Shantha Mahadevaiah and Mahesh Sangrithi (Turner laboratory), covering the same early developmental timepoint in which I had profiled DNA methylation. In opossum embryos, EGA occurs at E3.5; therefore, expression of methylation enzymes before this timepoint must derive from maternal deposition of transcripts in the oocyte.

I observed that the maintenance methyltransferase *DNMT1* and its partner protein *UHRF1* were highly expressed in the opossum oocyte, followed by a gradual drop in expression during the early cleavage stages (Figure 49a). *DNMT1* expression continued to fall throughout the developmental timecourse, reaching its lowest level at E7.5. In contrast, after reaching its lowest expression in the E4.5 embryo, *UHRF1* expression rose in E6.5 – 7.5 embryos (Figure 49a). In eutherian embryos, *DNMT1* and *UHRF1* are both expressed in the early embryo, but their protein products are largely restricted to the cytoplasm (Maenohara et al., 2017, Hirasawa et al., 2008). The de novo methyltransferases *DNMT3A* and *DNMT3B* were both found to be expressed in the oocyte. Both genes show a decrease in expression by E3.5, followed by a gradual rise in expression towards E7.5, suggestive of a transition from maternally-deposited transcript to zygotic expression (Figure 49b).

DNMT3L, which in eutherians is a non-catalytic protein partner of the de novo methyltransferases, showed negligible expression at all timepoints profiled in opossum (Figure 49b). In mouse oocytes, DNMT3L is expressed and required for the formation of correct methylation patterns (Kobayashi et al., 2012). Oocyte derived DNMT3L is degraded in early mouse embryos, and only becomes expressed from the zygotic genome in the E3.5 blastocyst, where it contributes to the formation the correct methylation landscape, but is not critical for successful development (Guenatri et al., 2013) . However,

this is not the case in human, where expression of DNMT3L is not detected in oocytes, only becoming expressed after fertilisation (Huntriss et al., 2004, Okae et al., 2014).

The *Tet* enzymes can stimulate active removal of DNA methylation by hydroxylation (Tahiliani et al., 2009, Ito et al., 2011). In the mouse, *Tet3* is highly expressed in the oocyte and zygotes, but rapidly declines in expression at the two-cell stage (Gu et al., 2011b, Iqbal et al., 2011). *Tet1* and *Tet2* become the predominant TETs expressed in the blastocyst. The action of TET1 and TET3 is required for correct gene expression and lineage specification in the pre-implantation embryo; double knockout of *Tet1/3* causes developmental arrest during the cleavage stages in a subset of embryos, and impaired formation of the ICM at the blastocyst stage (Kang et al., 2015). Of the three *TET* genes, only *TET1* was appreciably expressed in opossum early development. *TET1* was expressed in the oocyte, with expression levels falling slightly in E1.5 embryos, before rising to a peak of expression at E4.5. *TET1* expression then decreased gradually to reach levels similar to the oocyte by E7.5. The ‘fall-and-rise’ patterns of gene expression observed here suggest that *TET1* is maternally deposited in the oocyte and then degraded rapidly in the early embryo before onset of zygotic expression by E3.5.

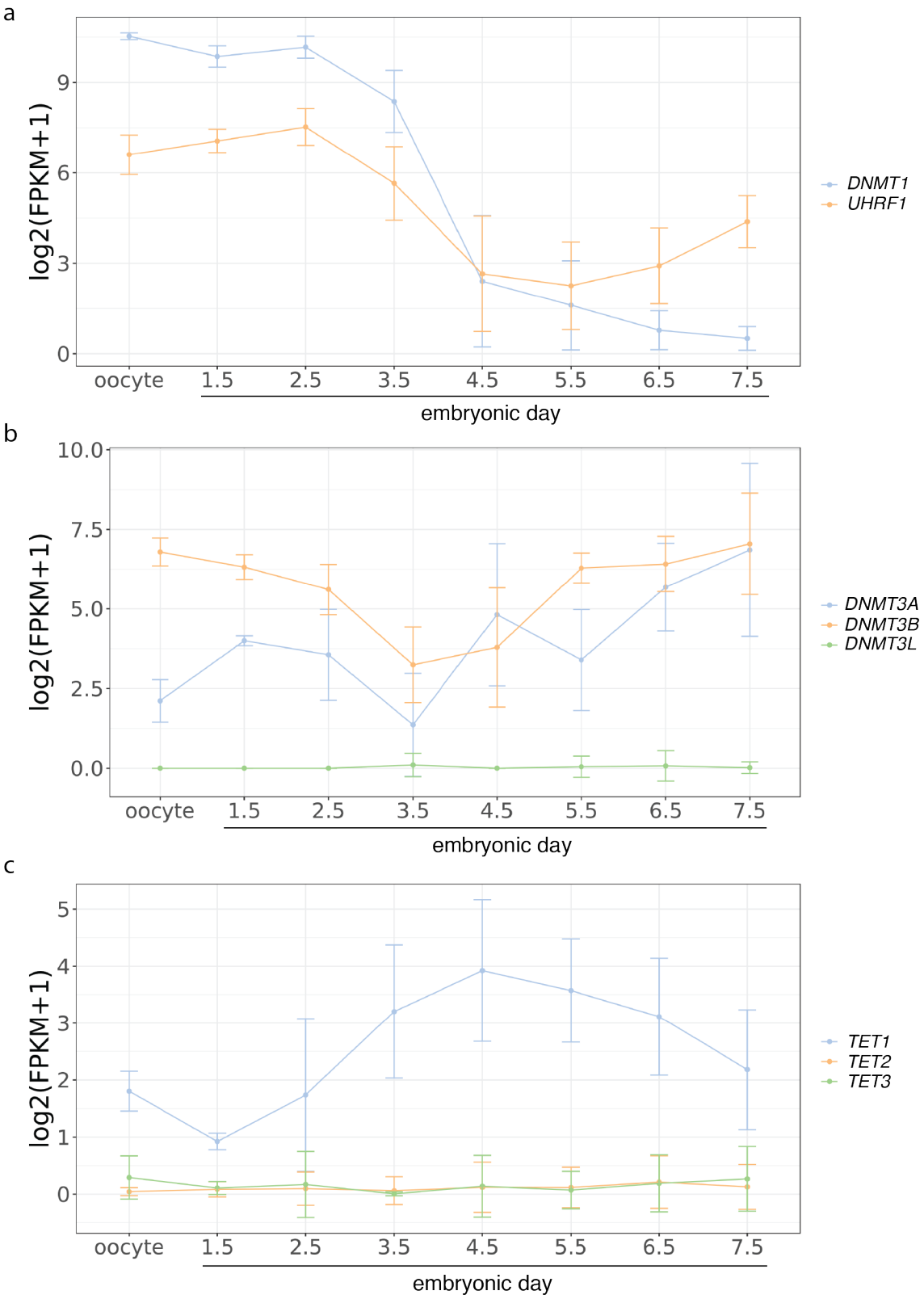

a
log2(FPKM+1)
0
3
6
9
oocyte
1.5
2.5
3.5
4.5
5.5
6.5
7.5
embryonic day
DNMT1
UHRF1
b
log2(FPKM+1)
0.0
2.5
5.0
7.5
10.0
oocyte
1.5
2.5
3.5
4.5
5.5
6.5
7.5
embryonic day
DNMT3A
DNMT3B
DNMT3L
c
log2(FPKM+1)
0
1
2
3
4
5
oocyte
1.5
2.5
3.5
4.5
5.5
6.5
7.5
embryonic day
TET1
TET2
TET3

Figure 49: Expression dynamics of methylation enzymes during opossum development. Expression dynamics of the opossum orthologues of enzymes known to be involved in DNA methylation. Data are mean log2+1 expression values from single cell RNA-seq (scRNA-seq) of opossum embryos. Error bars indicate standard deviation. scRNA-seq data were generated and processed by Shantha Mahadevaiah and Mahesh Sangrithi (Turner laboratory). **a**: Expression of the maintenance methyltransferase *DNMT1* and its partner enzyme *UHRF1*. **b**: Expression of the *de novo* methyltransferases *DNMT3A* and *DNMT3B*, and the non-catalytic partner protein *DNMT3L*. **c**: Expression of the enzymes *TET1*, *TET2* and *TET3*.

5.7 X-chromosome methylation during opossum development

In Chapters 3 and 4 I showed that the opossum Xi is broadly hypomethylated relative to the Xa and to autosomes in adult tissues. This finding raised the question of how and when this sex-specific difference arises. In this section, I explored this question by examining X-chromosome methylation across the timecourse of early male and female opossum development. In order to examine methylation of the X chromosome, I pooled individual BS-seq libraries by sex and timepoint, creating one *in silico* sample for each condition. Each timepoint pair was then filtered to include only CpG sites covered by at least 3 reads in each sex. Sperm and oocyte samples were also filtered in the same manner, to facilitate a direct comparison between the X-chromosome methylation of the gametes.

I compared the mean methylation level of autosomes, chrUn and the X chromosome between the gametes, and between males and females at each embryonic timepoint. (Figure 50a). Sperm and oocyte had very similar mean methylation levels on autosomes, while on chrUn oocytes were slightly hypomethylated (Figure 50a), reflecting the previously observed enrichment for sites of sperm:oocyte difference on this chromosome (Figure 45b). In embryos and adult tissues, the mean methylation of autosomes and chrUn was very similar between the sexes (Figure 50a). No difference in mean methylation level was observed on the X chromosome between sperm and oocytes (Figure 50a), suggesting that the inactive X is not inherited in a hypomethylated state from the paternal germline.

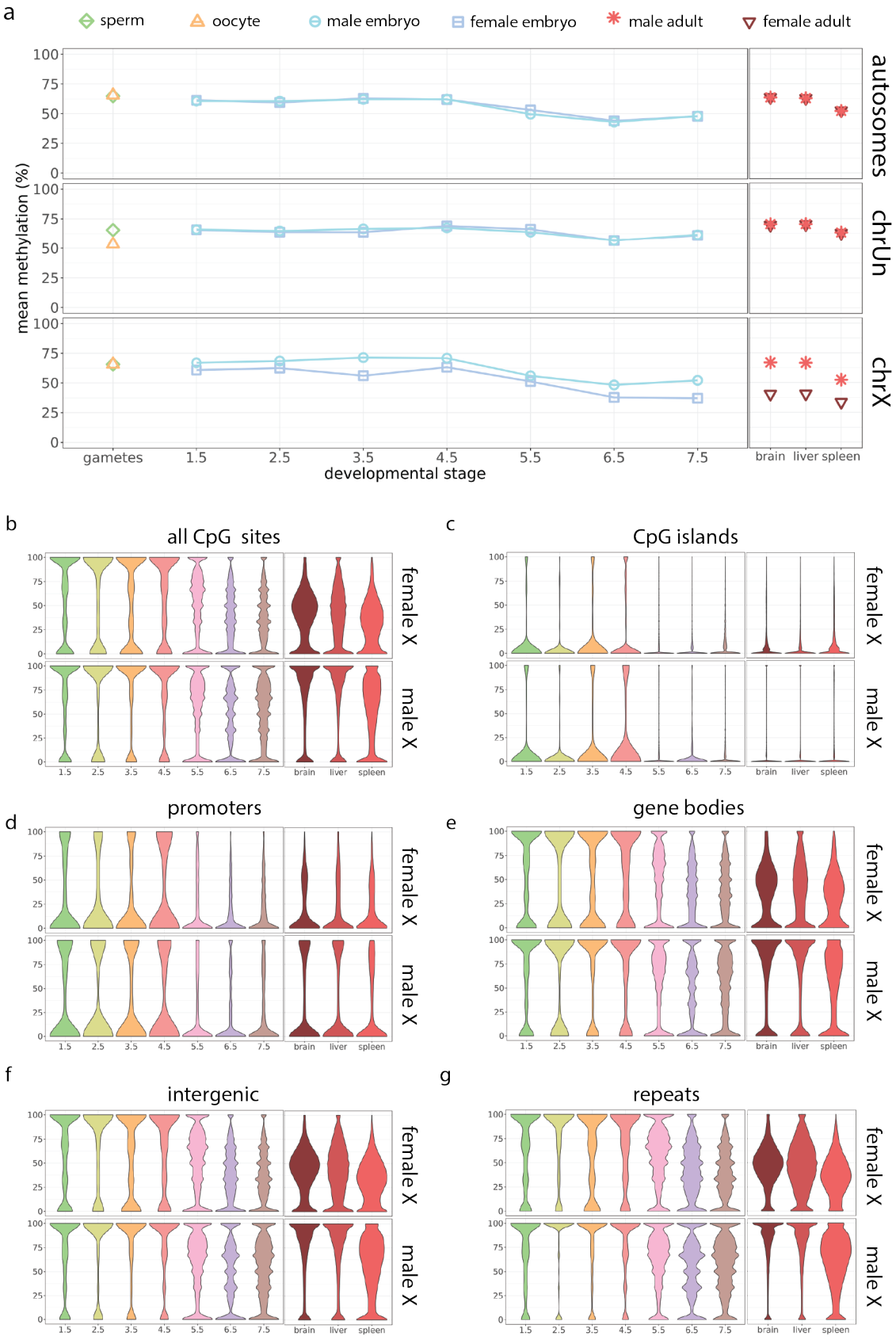

a
sperm
oocyte
male embryo
female embryo
male adult
female adult
mean methylation (%)
autosomes
chrUn
chrX
gametes
1.5
2.5
3.5
4.5
5.5
6.5
7.5
developmental stage
brain
liver
spleen
b
all CpG sites
c
CpG islands
d
promoters
e
gene bodies
f
intergenic
g
repeats
female X
male X

Figure 50: X chromosome DNA methylation dynamics during opossum development. Methylation levels of the X chromosome during opossum development. **a**: Global methylation level of autosomes, chrUn, and chrX in opossum gametes and male and female embryos from E1.5 – 7.5. Methylation was calculated as percentage of methylated CpGs/ covered CpGs per CpG site, and the mean of these values was taken as the global methylation level. $n_{gametes}$ = 82,872; $n_{1.5}$ = 644,050; $n_{2.5}$ = 157,315; $n_{3.5}$ = 320,549; $n_{4.5}$ = 195,102; $n_{5.5}$ = 8,185,325; ; $n_{6.5}$ = 9,047,553; $n_{7.5}$ = 10,698,945; n_{adult} = 4,445,874. **b-g**: Violin plots representing the overall methylation distribution of different genomic features on the X chromosome of male and female embryos from E1.5-7.5. y-axis represents methylation level (%). **b**: all CpG sites. **c**: CpG islands. **d**: promoters. **e**: gene bodies. **f**: intergenic regions. **g**: repeat elements.

At early embryonic stages, the mean X-chromosome methylation in males was similar to that on autosomes (Figure 50a). However, X-chromosome methylation in female embryos was lower than in males (Figure 50a). This difference was most pronounced at E3.5, and at E6.5-7.5, though no embryonic stage displayed a female:male difference as great as that seen in adult tissues (Figure 50a). To investigate these differences further, I examined overall distribution of methylation at X-linked CpG sites in both sexes (Figure 50b). The X chromosome displayed the same methylation dynamics as observed for the whole genome; namely a bimodal distribution skewed towards hypermethylation from E1.5-4.5, with a downwards shift in methylation level from E5.5-7.5 (Figure 50b). The lower mean methylation in E3.5 female embryos could be observed as a small population of CpG sites with intermediate methylation levels, within an overall distribution of hyper- and hypomethylated sites (Figure 50b). In contrast, from E5.5-7.5, both female and male embryos showed a depletion of highly methylated CpG sites, but this effect was amplified in female embryos (Figure 50b).

To understand whether the differences in methylation between male and female embryos derived from a particular genomic feature, I examined methylation distributions at different X-linked regions. I observed no difference between the sexes at CGIs and promoters, likely because these elements were consistently hypomethylated at all stages (Figure 50c-d). At gene bodies, intergenic regions, and repeat elements, I observed a methylation pattern reminiscent of the global trend; female embryos showed a higher number of regions with intermediate methylation level at early stages, and markedly fewer highly methylated regions at E6.5-7.5 (Figure 50e-g).

These results demonstrated that the female-specific X chromosome hypomethylation observed in adult tissues also occurs in opossum early embryos. XCI initiates at E3.5 in the opossum and silencing is complete by E4.5 (Shantha Mahadevaiah, Turner laboratory, unpublished). It is therefore possible that the female-specific hypomethylation relates to the onset of X-inactivation in female embryos.

In eutherian mammals, XCI is accompanied by hypermethylation of CGIs on the silenced chromosome. CGI hypermethylation is established by the *de novo* methyltransferase DNMT3B, and for the majority of CGIs this process relies on recruitment of the structural-maintenance of chromosomes protein *SMCHD1* to the Xi (Gendrel et al., 2012). In metatherians, hypermethylation of Xi CGIs has been suggested not to occur (Cooper et al., 1993, Samollow et al., 1989, Migeon et al., 1989, Wang et al., 2014b, Waters et al., 2018), an observation confirmed by the allele-specific analyses I conducted in Chapter 4. Given the absence of CGI hypermethylation in marsupials, it was interesting to consider the expression patterns of *DNMT3B* and *SMCHD1* in relation to the timing of opossum XCI. I examined the expression of these genes using opossum embryo scRNA-seq data generated by Shantha Mahadevaiah and Mahesh Sangrithi (Turner laboratory, unpublished). *DNMT3B* and *SMCHD1* were both expressed in the opossum oocyte and during early development, including during the onset of XCI at E3.5, and at subsequent stages (Figure 51). These results suggest that the absence of CGI methylation on the opossum Xi is not attributable to a lack of expression of the factors which facilitate CGI methylation in eutherians.

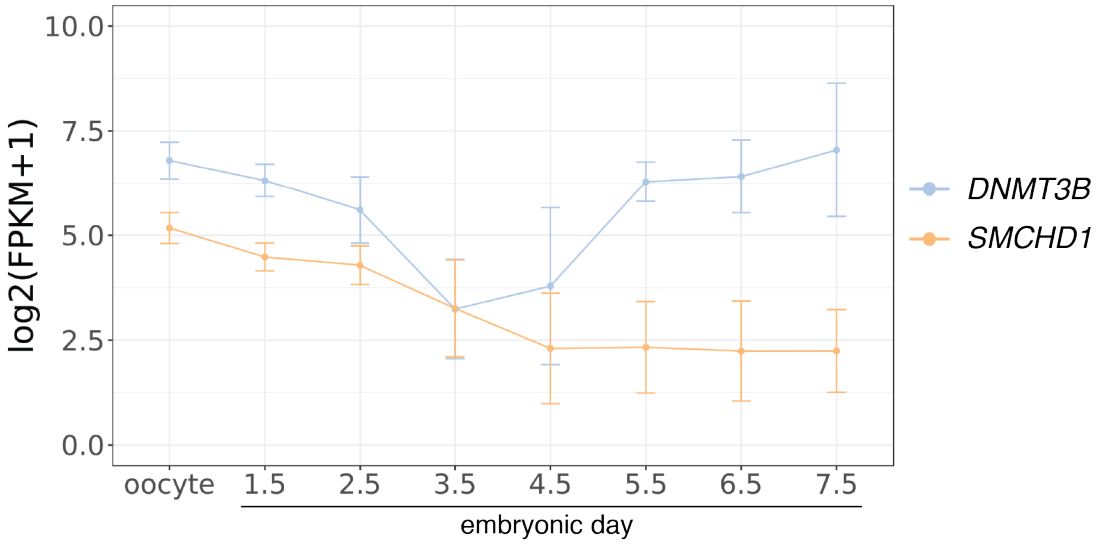


Figure 51: Expression of *SMCHD1* and *DNMT3B* during opossum development. Expression dynamics of the opossum orthologues of *SMCHD1* and *DNMT3B* across a timeline of opossum early development. Data are mean log2+1 expression values from single cell RNA-seq (scRNA-seq) of opossum embryos. *DNMT3B* is replotted from Figure 49b. Error bars indicate standard deviation. scRNA-seq data were generated and processed by Shantha Mahadevaiah and Mahesh Sangrithi (Turner laboratory).

Metatherian XCI is imprinted such that the paternal X chromosome is always the silenced chromosome. The molecular nature of the imprint that establishes this silencing is not known, but one candidate is differential methylation of the promoters of *RSX*/*XSR*. I therefore examined methylation patterns at the *RSX* and *XSR* locus.

Initially I examined methylation at the *RSX* locus in adult tissues, using the BS-seq data generated for brain, liver, and spleen in Chapter 4 (Figure 52a). Biological triplicate libraries from each tissue were pooled according to sex, and filtered to include only CpGs covered at a minimum read depth of three. As a control, I examined methylation at the autosomal *RUNX1* locus, and did not observe any obvious sex-specific differences (Figure 52b). Examination of the *RSX* locus revealed that the promoter/CGI region upstream of *RSX* is highly methylated in all male tissues examined but methylated at an intermediate level in all female tissues (Figure 52a). This result confirms that the *RSX* promoter DMR found in foetal brain and extra-embryonic membranes (Wang et al., 2014b) also occurs in adult tissues, and suggests that if the Xa is hypermethylated similar to the male, then the Xi is hypomethylated at this locus. Inspection of the promoter region of *XSR* did not reveal any obvious sex differences but did reveal a region of hypomethylation present in all samples (Figure 52a).

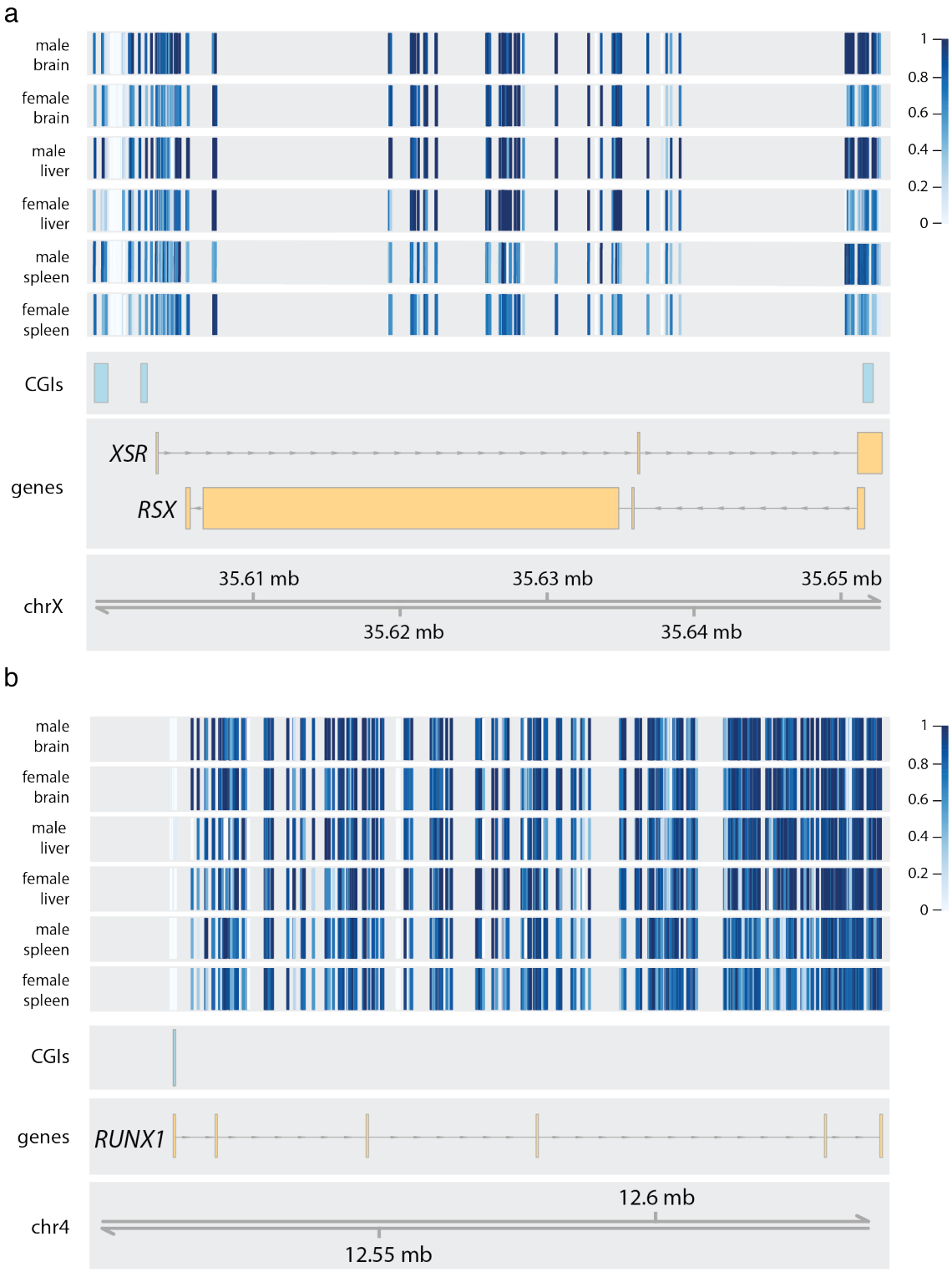

a
male brain
female brain
male liver
female liver
male spleen
female spleen
1
0.8
0.6
0.4
0.2
0
CGIs
genes
XSR
RSX
chrX
35.61 mb
35.63 mb
35.65 mb
35.62 mb
35.64 mb
b
male brain
female brain
male liver
female liver
male spleen
female spleen
1
0.8
0.6
0.4
0.2
0
CGIs
genes
RUNX1
chr4
12.6 mb
12.55 mb

Figure 52: DNA methylation at the *RSX* locus in adult opossums.
DNA methylation level at individual CpG sites in adult opossum brain, liver, and spleen. Data are adult BS-seq generated in Chapter 4. Three biological replicates were pooled per condition, and CpG sites were filtered to a minimum coverage of three reads. Methylation level was calculated as the proportion of methylated Cs/ covered Cs per site. Each tile represents one CpG site, with high methylation level in blue and low methylation level in white. Gene features, CpG islands, and chromosomal position are indicated below. **a**: Methylation at the *RSX/XSR* locus. CpG islands in this plot were predicted by Fanny Decarpentrie (Turner laboratory) using the DBCAT CpG island finder (Kuo et al., 2011). **b**: Methylation at the *RUNX1* autosomal locus is shown for comparison. CpG islands in this plot are as annotated by UCSC.

I next considered methylation of the *RSX* locus in opossum gametes (Figure 53). For sperm libraries, the sparse nature of low-input BS-seq data resulted in very few covered CpG sites and precluded interpretation of promoter methylation state (Figure 53a). The sparsity of sperm data was also apparent at the autosomal *RUNX1* locus (Figure 53b). However, in oocytes sufficient CpG sites were covered to reveal that the *RSX* promoter is highly methylated, while the *XSR* promoter is unmethylated (Figure 53a).

Subsequently, I examined the *RSX* locus in male and female embryos (Figure 54a). At E1.5, male embryos showed high methylation at the *RSX* promoter, and low methylation at the *XSR* promoter, suggesting that the X chromosome inherited from the oocyte retains its methylation patterns at the *RSX* locus following fertilisation. Sparse CpG coverage in samples from E1.5 to E4.5 precluded interpretation of the methylation state of the *RSX* and *XSR* promoters at these timepoints.

In samples from E5.5 to 7.5, coverage improved sufficiently to reveal that the *RSX* promoter carries different methylation levels in male and female embryos (Figure 54a), in a manner reminiscent of the pattern in adult tissues (Figure 54a). Likewise, the region within the *XSR* promoter that was hypomethylated in adults was similarly hypomethylated in E5.5 – 7.5 embryos. By comparison, at the *RUNX1* locus there was no obvious sex-specific difference in DNA methylation (E7.5 embryos shown as an example; Figure 54b).

If the methylation states inherited from the gametes are maintained throughout development, these results imply that the active X carries a highly methylated *RSX*

promoter, while the same locus on the inactive X is hypomethylated. Conversely, the promoter of *XSR* would appear to be unmethylated on both the active and inactive X chromosome. However, it not possible to rule out local reprogramming of the *RSX* and *XSR* promoter regions during the E1.5 – 4.5 period which was unresolved in this analysis. The sequencing of further opossum embryos BS-seq libraries to attain broader coverage of the locus would permit this question to be resolved.

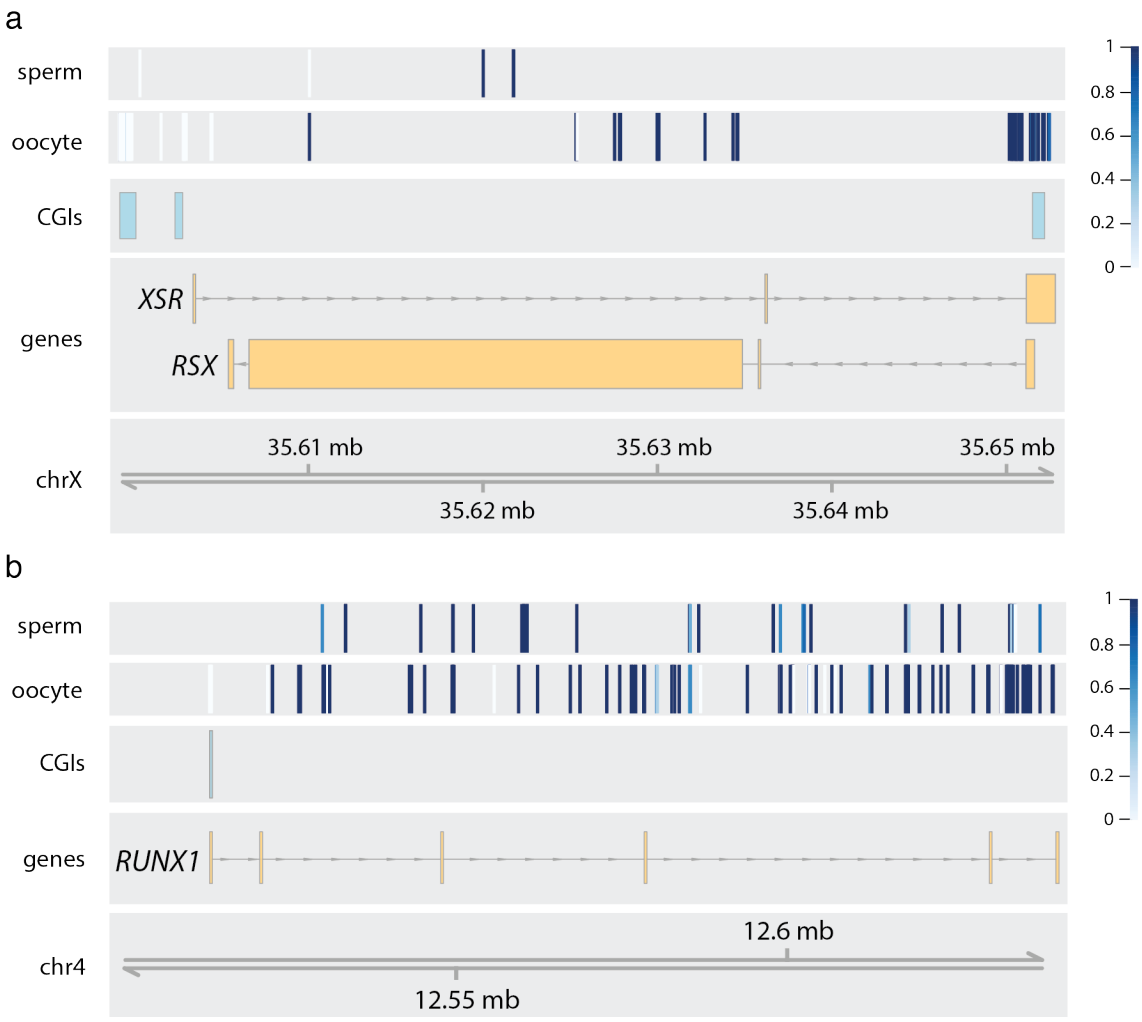


Figure 53: DNA methylation at the *RSX* locus in opossum gametes. DNA methylation level at individual CpG sites in opossum sperm and oocytes. Data from individual libraries were pooled and subjected to a minimum coverage filter of 3 reads per site. Methylation level was calculated as the proportion of methylated Cs/ covered Cs per site. Each tile represents one CpG site, with high methylation level in blue and low methylation level in white. Gene features, CpG islands, and chromosomal position are indicated below. **a**: Methylation at the *RSX/XSR* locus. CpG islands in this plot were predicted by Fanny Decarpentrie (Turner laboratory) using the DBCAT CpG island finder (Kuo et al., 2011). **b**: Methylation at the *RUNX1* autosomal locus is shown for comparison. CpG islands in this plot are as annotated by UCSC.

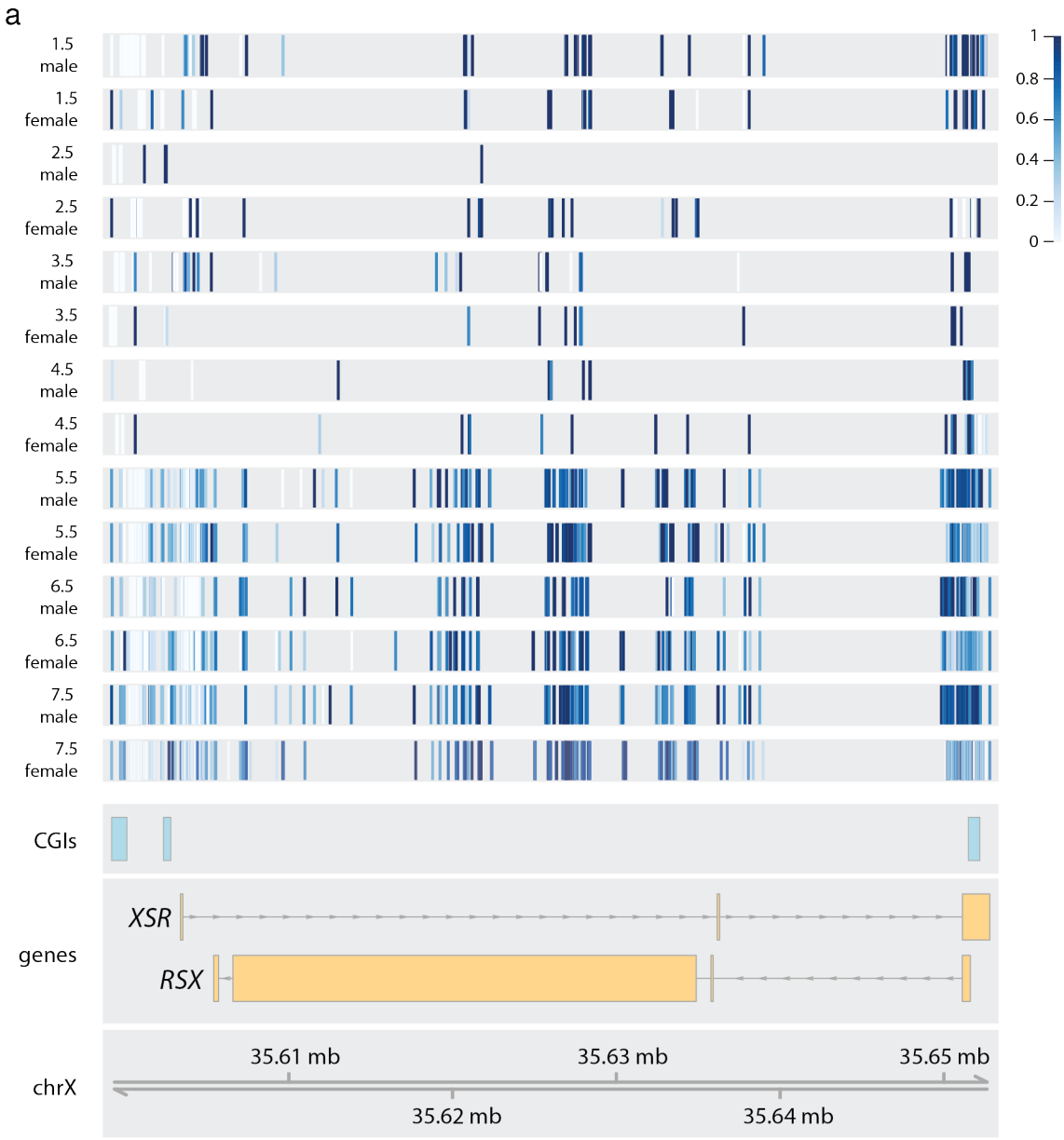

a
1.5 male
1.5 female
2.5 male
2.5 female
3.5 male
3.5 female
4.5 male
4.5 female
5.5 male
5.5 female
6.5 male
6.5 female
7.5 male
7.5 female
1
0.8
0.6
0.4
0.2
0
CGIs
genes
XSR
RSX
chrX
35.61 mb
35.63 mb
35.65 mb
35.62 mb
35.64 mb
b
7.5 male
7.5 female
CGIs
genes
RUNX1
chr4
12.6 mb
12.55 mb

Figure 54: DNA methylation at the *RSX* locus in opossum embryos. DNA methylation level at individual CpG sites in opossum male and female embryos. Data from individual libraries were pooled and subjected to a minimum coverage filter of 3 reads per site. Methylation level was calculated as the proportion of methylated Cs/ covered Cs per site. Each tile represents one CpG site, with high methylation level in blue and low methylation level in white. Gene features, CpG islands, and chromosomal position are indicated below. **a**: Methylation at the *RSX/XSR* locus. CpG islands in this plot were predicted by Fanny Decarpentrie (Turner laboratory) using the DBCAT CpG island finder (Kuo et al., 2011). **b**: Methylation at the RUNX1 autosomal locus in E7.5 embryos is shown for comparison. CpG islands in this plot are as annotated by UCSC.

5.8 Discussion

In this chapter, I analysed DNA methylation over a timeline of opossum early development. Initially I focussed on evaluation of possible low-input bisulfite sequencing methods using mouse embryos and low amounts of purified gDNA as optimisation controls. I then generated low-input BS-seq libraries from a developmental timecourse encompassing sperm, oocytes, and embryos from 1.5 to 7.5 days post-coitum. This timecourse encompassed the E3.5-4.5 period during which XCI occurs in opossum, and in addition included the emergence of the extra-embryonic and embryonic lineages at days 6.5-7.5.

The BS-seq method is reported to capture up to 50% of genomic CpGs when applied to mouse single-cell samples. In the libraries generated here, coverage correlated with amount of input material. Libraries prepared from later embryonic stages attained coverages approaching 50% of CpGs, while timepoints earlier than E5.5 did not reach this level. The opossum genome is 3.6 Gb, approximately 1 Gb larger than the mouse, and therefore it would be necessary to sequence opossum samples more deeply than mouse to attain coverage equivalent coverage. The low-input BS-seq method suffers from relatively low mapping rates due to chimeric reads that arise during cycles of randomly primed strand synthesis. A recently developed mapping method attempts to address the low mapping rates, by performing local realignment of unmapped reads (Wu et al., 2019). Future work applying this method to the data generated here might attain even higher genomic coverage.

Examination of the global methylation distribution of opossum gametes and embryos suggested that the opossum does not undergo a large-scale demethylation following fertilisation, nor a sustained phase of global hypomethylation during early embryogenesis. Most CpG sites profiled did not change between consecutive timepoints. The greatest difference in methylation was seen between sperm and oocyte, a pattern that is also observed in the gametes of eutherian mammals and of zebrafish (Potok et al., 2013, Jiang et al., 2013, Smallwood et al., 2011, Smith et al., 2012). Differentially methylated sites between the gametes reflected methylation at genic regions in oocytes, similar to observations in mouse oocytes (Veselovska et al., 2015). I also observed at a bulk level that opossum oocytes harbour a slightly elevated non-CpG methylation rate compared to embryos and adult tissues, similar to the trend observed in mouse oocytes (Tomizawa et al., 2011). In future, it would be interesting to explore the genomic patterns of non-CpG methylation in opossum oocytes in greater detail.

The bulk methylation level of E1.5 embryos was intermediate between sperm and oocyte levels. This result suggests that global methylation level at E1.5 represents the average of the sperm and oocyte methylomes, implying that there is no rapid demethylation of the paternal genome by the two-cell stage as seen in the mouse (Smith et al., 2012). This finding was supported by 5mC immunostaining of opossum zygotes, which revealed the two pronuclei to harbour similar methylation levels. However, a caveat precludes making this a conclusive interpretation. It was not possible to identify the polar bodies in E1.5 embryos, and therefore they were not removed prior to collection for sequencing. A detailed zygotic staging has not yet been performed for the opossum, and it is therefore unknown at what point polar body breakdown occurs. Experiments in human embryos have shown that the polar bodies retain similar methylation patterns to the maternally contributed pronuclear genome (Guo et al., 2014b). It is therefore possible that the methylation level seen in E1.5 opossum embryo samples contained a contribution from the polar body genome and was therefore not the strict average of the sperm and oocyte, but was instead skewed towards the oocyte-contributed methylation pattern. With the assumption that the polar body genome would display a slightly hypomethylated genome similar to the oocyte, this would suggest that the estimated bulk methylation level at E1.5

could be an underestimate. In future, this question will be resolved by low- input BS-seq of zygotes from which the polar bodies have been removed.

The observation that most loci did not change methylation between consecutive timepoints suggests that in opossums, totipotency is established via local reprogramming of some genomic regions, as observed in the zebrafish, rather than by global hypomethylation followed by remethylation as seen in eutherians mammals (Potok et al., 2013, Jiang et al., 2013, Smith et al., 2012, Smith et al., 2014, Guo et al., 2014b). I observed a slight drop in methylation in embryos between E5.5-7.5. The opossum extra-embryonic tissues were shown to be hypomethylated relative to the embryonic lineage in embryos at the trilaminar blastocyst stage (Stevens et al., 1988). I therefore hypothesise that the decrease in methylation I observed in blastocyst stage embryos is due to the higher representation of cells from the extra-embryonic lineage at this timepoint (Figure 41a). To confirm this hypothesis, future work could perform BS-seq of separated embryonic and extraembryonic compartments, either by disaggregation of embryos to single cells, or by dissection of the embryonic disk from the surrounding extra-embryonic tissue. To validate these bulk methylation findings using an independent method, future work could profile the bulk methylation levels of key timepoints in opossum development using ultra-sensitive mass spectrometry (Amouroux et al., 2016).

Examination of expression of the opossum orthologues of factors involved in DNA methylation revealed that the maintenance methylation enzyme *DNMT1* and its partner protein *UHRF1* are expressed in opossum oocytes and embryos. Presumably the availability of the maintenance methylation machinery accounts for the stability of methylation state observed in the opossum embryo and may be one reason there was not a large global methylation reprogramming. However, in the mouse embryo, *UHRF1* and *DNMT1* are present at the transcript level, but their proteins are almost entirely located in the cytoplasm (Maenohara et al., 2017, Hirasawa et al., 2008). It would be interesting to confirm the subcellular localisation of the opossum DNMT1 and UHRF1 orthologues during embryonic development. It is interesting to note that in mouse oocytes and embryonic stem cells, the cytoplasmic localisation of UHRF1 and DNMT1 is mediated by STELLA, a gene for which there is no annotated marsupial orthologue (Li et al., 2018,

Mulholland et al., 2020). The *de novo* methyltransferases *DNMT3A* and *DNMT3B* were both expressed in opossum oocytes and early embryos, and are perhaps involved in adding methylation at the few regions I observed to gain methylation during development. *DNMT3L*, which in eutherians is a non-catalytic partner protein involved in determining the activity of the *de novo* methyltransferases, is not expressed to appreciable levels in the opossum embryo.

In eutherians, *TET3* is highly expressed in the oocyte and zygote, and was suggested to be responsible for the active removal of methylation from the paternal pronucleus (Gu et al., 2011b, Iqbal et al., 2011, Wossidlo et al., 2011). However, more recent evidence suggests that *TET3* is not the enzyme responsible for initial active demethylation upon fertilisation, but functions to maintain the demethylated state at some loci in the early embryo (Salvaing et al., 2012, Santos et al., 2013, Peat et al., 2014, Wang et al., 2014a, Guo et al., 2014a, Shen et al., 2014, Amouroux et al., 2016). All three TET proteins are involved in regulating gene expression and the balance between pluripotency and differentiation in later stages of eutherian development (Dawlaty et al., 2013, Dawlaty et al., 2014, Kang et al., 2015, Zhang et al., 2018). In opossums, *TET2* and *TET3* were not expressed to appreciable levels during early development, while *TET1* expression was present in the oocyte, dropped in the E1.5 embryo, and rose again in E2.5- 7.5 embryos. This suggests that the developmentally regulated roles of *TET* enzymes differ between the eutherian and metatherian lineages, and may account for the absence of rapid demethylation of the paternal pronucleus in opossums.

It has been hypothesised that the phase of sustained hypomethylation in the preimplantation eutherian embryo relates to the uniquely mammalian requirement for early activation of the embryonic genome (Hackett and Surani, 2013). Early EGA is thought to have evolved in mammals to permit the differential gene expression required to specify the embryonic and extra-embryonic lineages, a requirement not present in zebrafish, where genome-wide reprogramming does not occur (Hackett and Surani, 2013). Global demethylation is thought necessary for the early embryonic genome to reach a totipotent state in time for this differential expression. If embryonic hypomethylation is an absolute requirement for the specification of extra-embryonic

lineages, it would be expected to also occur in marsupials, where specification of this lineage also occurs. The results presented here suggest that genome-wide reprogramming in the early embryo is a eutherian phenomenon, and suggest that opossum embryos attain totipotency via local reprogramming of methylation rather than large global changes.

In addition to the global study of methylation dynamics undertaken here, I utilised the opossum low-input BS-seq dataset to interrogate DNA methylation patterns of the X chromosome during the establishment of XCI. In adult opossum tissues, it is well established that CGI regions on the Xi are not hypermethylated. I observed that *SMCHD1* and *DNMT3B*, which facilitate Xi CGI methylation in the eutherian embryo, are expressed during the E3.5-4.5 period of opossum XCI establishment. This suggested that the lack of CGI methylation was not due to absence of the factors that perform this function in eutherians. Even if *SMCHD1* does not facilitate methylation of CGIs on the Xi, it could play other roles, perhaps in the establishment of other repressive chromatin marks to ensure the silencing of the X chromosome. It would be interesting to profile 3D structure of the opossum Xi, for example by Hi-C, though this would be challenging given the low SNP density that is currently achievable in opossum crosses.

In eutherians, Xi CGI hypermethylation is a late step in XCI, established during the global remethylation of the genome during exit from pluripotency as 'a locking-in' of the silent state that has already been established by chromatin changes. The apparent absence in opossum embryos of a global demethylation and remethylation perhaps explains the absence of CGI hypermethylation on the Xi; there may not be a phase during which *de novo* methylation can access Xi CGIs and change their methylation state. However, in adult opossum tissues I observed that the entire Xi harboured low methylation levels. In this chapter, I observed that the X chromosome in female embryos displayed a lower methylation level than males. This effect was evident at the time of XCI at E3.5, and became more pronounced in later stages. However, minor hypomethylation of the female X chromosome was evident even in the preceding 1.5-E2.5 timepoints, raising the possibility that XCI-associated methylation changes precede the onset of XCI. However, the X chromosome in sperm and oocytes carried equivalent methylation levels. These findings raise the question of how methylation is lost from the Xi. The slow rate of the

loss in methylation is suggestive of passive demethylation by exclusion of the maintenance methylation machinery. The opossum X chromosome is late replicating, so it is feasible that temporal regulation of the nuclear localisation of maintenance methylation machinery could exclude the Xi. In future, immunostaining experiments to establish the localisation of DNMT1 and UHRF1 during the cell cycle, and their colocalization with the Xi, could shed light on this question. It is also possible that loss of methylation from the Xi is partly or fully enacted by active demethylation, for example via the action of *TET1,* which is expressed during the embryonic period.

Examination of methylation at the *RSX* locus in opossum oocytes revealed that the *RSX* promoter is highly methylated, while the *XSR* promoter is unmethylated, but low-resolution data prevented any similar conclusion being drawn from sperm and early embryos. In later stage embryos, a pattern reminiscent of adult tissues was apparent, with high methylation at the *RSX* promoter in males, and intermediate methylation in females. The *XSR* promoter did not appear to carry different methylation between the sexes at any timepoints. Potentially, transcription across the *XSR* locus during oogenesis establishes DNA methylation at the *RSX* promoter, thereby preventing *RSX* expression in *cis* from the maternal X chromosome in the embryo. Continued expression of *XSR* in the embryo could perhaps be acting as a backstop mechanism to ensure the maintenance of this imprint. This system in some ways resemble the observation that methylation of the *Xist* promoter occurs via the transcription of *Tsix* (Ohhata et al., 2008). The intermediate methylation level at the *RSX* promoter in female embryos is suggestive of this region being inherited in an unmethylated state from the sperm, thereby setting up a reciprocal imprint to permit *RSX* expression from the Xi. Interestingly, a sex-specific difference in methylation was not observed at the *XSR* promoter in female embryos, suggesting that this region may also be unmethylated in sperm. Further examination of the *RSX* locus in sperm sample with improved coverage would verify the interpretation of these findings.

Chapter 6. Discussion

DNA methylation is an important epigenetic mark with roles in the regulation of gene expression and the silencing of transposable elements. In eutherian mammals, DNA methylation is required for successful embryonic development (Li et al., 1992b, Okano et al., 1999). During early development of eutherian mammals, the genome is subject to global demethylation, such that preimplantation development is completed while the genome harbours an atypically low level of methylation (Mayer et al., 2000, Oswald et al., 2000, Smith et al., 2012). Upon exit from pluripotency, methylation is returned to somatic-like levels. The global removal of DNA methylation is thought to be necessary for the embryo to attain totipotency.

Silencing of one X chromosome in female eutherian achieves dosage compensation for X-linked gene products between male and female mammals (Lyon, 1961). Eutherian X inactivation is accomplished in the early embryo via the action of the lncRNA *XIST* and a suite of repressive epigenetic changes. The final step in the silencing of the Xi in eutherian mammals is the establishment of repressive methylation at the promoters of silent genes, a process that occurs concomitantly with the global remethylation of the genome during the exit from pluripotency (Sado et al., 2000, Auclair et al., 2014).

The requirement for X-dosage compensation arose with the evolution of the mammalian XY sex chromosomes and therefore this particular role for DNA methylation is unique to mammals. Similarly, it has been shown that non-mammalian vertebrates do not undergo a genome-wide demethylation event in the early embryo, as exemplified by studies of the zebrafish (Potok et al., 2013, Jiang et al., 2013). Evidently, DNA methylation holds dynamic and important roles during eutherian embryogenesis. Understanding the role of DNA methylation in the early development and X chromosome inactivation of metatherian mammals can therefore shed light on the evolution of these processes in the mammalian lineage. In this thesis I have examined DNA methylation in the early embryos and adults of a metatherian mammal, the opossum *M. domestica*. Here I will consider the significance of my results to some aspects of the evolution of genome-wide reprogramming and X chromosome inactivation.

6.1 Genome wide reprogramming

In this thesis, I observed that DNA methylation levels in opossum embryos were not dramatically reprogrammed following fertilisation (Chapter 5; Figure 55a). Global methylation levels were broadly stable for the first four days of embryonic development, and the paternal genome was not observed to undergo dramatic demethylation in the zygote. Subsequently, from E5.5-E7.5, methylation levels declined slightly. At none of the stages profiled did methylation levels decrease to the extent seen in the preimplantation embryos of eutherian mammals (Smith et al., 2012, Smith et al., 2014, Guo et al., 2014b). Global methylation measurements using immunostaining and bisulfite sequencing of targeted loci have previously demonstrated that DNA methylation is reprogrammed in the germline of the tammar wallaby (Suzuki et al., 2013, Ishihara et al., 2019). Taken together, these results suggests that the evolutionary acquisition of genome-wide reprogramming could have occurred in two steps; arising first in the germline in the common ancestor of therian mammals, and subsequently in the post-fertilisation embryo in the eutherian lineage. However, this conclusion is an inference based on study of only two metatherian species; the tammar wallaby germline (Suzuki et al., 2013, Ishihara et al., 2019), and the opossum early embryo (this work). To confirm this interpretation, it will be important to assay DNA methylation at base-resolution in the germline and early embryos of diverse metatherian species.

Regardless, is interesting to consider potential reasons for the evolution of GWR, and why this may have occurred differently between the metatherian and eutherian lineages. The reasons for GWR can be thought of both in terms of proximate causes, i.e. the mechanism involved, and ultimate causes, i.e. the evolutionary advantages of acquiring such a mechanism.

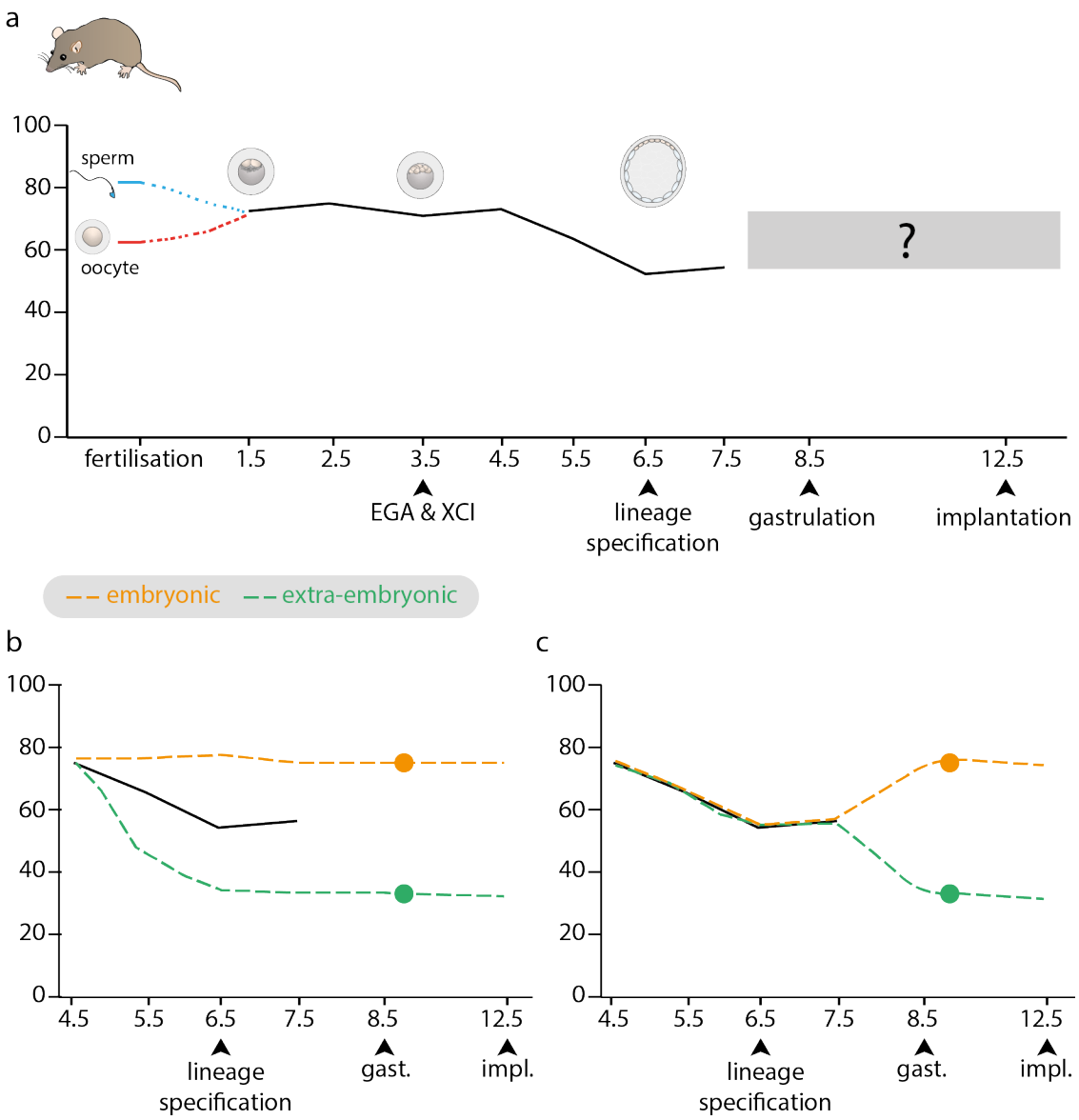


Figure 55: Methylation dynamics during early development of the opossum. Schematic representing methylation levels during opossum preimplantation development. **a**: Representation of the median methylation levels in sperm, oocytes, and embryos from E1.5 – 7.5 as established in this thesis. Dashed red and blue lines symbolise that the methylation level at E1.5 could be the average of the two gametes. Methylation levels after E7.5 have not yet been examined by base-resolution methods (grey box). **b-c**: Two possible methylation scenarios in the embryonic and extra-embryonic lineages. Extra-embryonic lineages were previously shown to be hypomethylated relative to embryonic lineages in the opossum embryo (green and orange circles). The depicted level of methylation is arbitrary and serves to illustrate the Southern blot findings (Stevens et al., 1988). **b**: Scenario I: The embryonic compartment may retain high levels of methylation, while the extra-embryonic compartment is demethylated. **c**: Scenario II: Both embryonic and extra-embryonic lineages may lose methylation, followed by subsequent remethylation of the embryonic lineage. Gast: gastrulation, imp: implantation.

6.1.1 **Reprogramming in the germline**

The presence of germline GWR in all therians is consistent with the idea that an original impetus for the methylation removal was the resetting of imprinted alleles (Reik and Walter, 2001a). Imprinting evolved in the therian ancestor, and therefore germline resetting of imprints must have arisen simultaneously. However, imprint erasure does not explain the genome-wide loss of methylation. It has been suggested that global removal of methylation in the germline functions to prevent the transmission of epimutations to the next generation (Reik et al., 2001). Perhaps the expression of demethylation machinery in the germline arose to facilitate imprint erasure, and was subsequently co-opted for a wider role, such as the removal of epimutations. If germline reprogramming evolved in the common therian ancestor, then it is possible that the mechanisms of DNA methylation removal acting in the germline are shared between metatherians and eutherians. In future it would be interesting to test whether factors implicated in eutherian germline reprogramming have similar roles in the metatherian germline.

6.1.2 **Reprogramming in the embryo**

6.1.2.1 Proximate causes

In contrast to the germline, mechanisms involved in early embryo GWR may have been acquired or co-opted uniquely in the eutherian lineage. In eutherians, several mechanisms are implicated in the removal of DNA methylation in the zygote, including base excision repair, TET-mediated hydroxymethylation, and passive demethylation via replicative dilution. An exact understanding of how these activities function together is still emerging.

The rapid demethylation of the paternal genome at fertilisation in eutherian mammals has been linked to the DNA repair processes carried out by the oocyte after fertilisation (Hajkova et al., 2010, Wossidlo et al., 2010, Santos et al., 2013). Sperm are subject to DNA damage due to reactive oxygen species (ROS) and shear forces in the female reproductive tract, but are also exposed to ROS due to the packaging of DNA onto protamines (Champroux et al., 2016). Epididymal ROS exposure induces the formation of disulphide bridges between cysteines on protamines, ensuring the extreme

condensation of the sperm nucleus (Champroux et al., 2016, Marushige and Marushige, 1975, Balhorn, 1982). In eutherian mammals, paternal genomic integrity is ensured by factors in the zygote cytoplasm that enact a programme of DNA repair on the sperm pronucleus immediately after fertilisation. The replacement of methylated cytosine with the unmethylated form during DNA repair contributes to the rapid loss of methylation from the paternal pronucleus.

The retention of paternal methylation at fertilisation in metatherians could indicate that DNA-repair processes do not occur in the same manner in the metatherian zygote. A hint in this direction comes from the observation that metatherian protamines do not contain cysteine and therefore cannot form disulphide bridges in the compacted sperm nucleus (Retief et al., 1995). This observation predicts that metatherian sperm would be less robustly condensed than eutherian sperm, which was confirmed by observations of metatherian sperm instability (Cummins, 1980), and is reflected in my observation that opossum sperm are more vulnerable to detergent lysis than eutherian sperm. The absence of cysteines in metatherian protamines leads to the speculation that metatherian sperm might not suffer such extensive oxidative damage as eutherian sperm. Therefore, it is possible that requirements for DNA repair in the zygote could have evolved differently in metatherian and eutherian lineages. It is noteworthy that zebrafish sperm are packaged by histones rather than protamines, and that in this species paternal DNA methylation is retained after fertilisation. Indeed, in various non-eutherian mammalian clades, such as insects, birds, reptiles and bony fish, protamines lack cysteine (Balhorn, 2007). To test these ideas, markers of DNA damage should be examined in metatherian zygotes, alongside expression patterns of DNA repair enzymes. It would also be interesting to correlate the methylation state of paternal pronuclei after fertilisation with sperm packaging strategy across a wide range of organisms.

In eutherian mammals, there is further localised demethylation by TET3 in both the maternal and paternal pronuclei (Gu et al., 2011b, Iqbal et al., 2011, Wossidlo et al., 2011). *TET3* is expressed in the early eutherian embryo but decreases during the cleavage stages and is replaced by *TET1* and *TET2* in the ICM of the blastocyst (Gu et al., 2011b, Iqbal et al., 2011, Wossidlo et al., 2011, Ruzov et al., 2011). In contrast, in opossums

TET1 is expressed in the early embryo and the blastocyst and *TET3* expression is negligible during early development. Therefore there is still the potential for TET-mediated demethylation in opossum embryos, but it is clear that the role acquired by *TET3* in eutherian early development is absent in metatherians. It is interesting to speculate that opossums may move directly to a programme of TET1 usage resembling the pluripotent cells of the eutherian blastocyst, while eutherians first enact a different, TET3-driven programme.

In addition to mechanisms of active removal, DNA methylation in the eutherian embryo is lost by replicative dilution due to the exclusion of UHRF1 and DNMT1 from the nucleus (Howell et al., 2001). The maintenance of high levels of methylation in the opossum embryo suggests that UHRF1 and DNMT1 must access the nucleus in this species. However, in eutherians, a small amount of DNMT1 is present in the nucleus, and is responsible for maintaining methylation at imprinted genes in the context of global hypomethylation. The oocyte-produced form, DNMT1O, is responsible for imprint maintenance for only one cell cycle, at the eight-cell stage; DNMT1S maintains imprints at other stages of preimplantation development (Kurihara et al., 2008, Hirasawa et al., 2008). Intriguingly, DNMT1O is conserved in the opossum (Ding et al., 2003). Given that drastic demethylation is not apparent in the opossum embryo, it is unclear whether a specialised role for DNMT1O in the maintenance of imprints would be required. Examination of the intracellular localisation patterns of DNMT1S, DNMT1O, and UHRF1 will help to clarify the role of maintenance methylation enzymes in the methylation landscape of the early opossum embryo.

In eutherians, STELLA has been implicated in protecting imprinted regions and the overall maternal genome from active demethylation (Nakamura et al., 2007, Nakamura et al., 2012, Bian and Yu, 2014). However, STELLA also protects the oocyte methylome from aberrant methylation accumulation by sequestering UHRF1 in the cytoplasm (Li et al., 2018), and maintains hypomethylation in ESCs via the same mechanism (Mulholland et al., 2020). STELLA therefore holds opposing functions; promoting or preventing methylation depending on context. In metatherians, there is no *STELLA* orthologue (Ensembl release 99, Cunningham et al., 2019). It is therefore tempting to speculate that

global demethylation arose in eutherian mammals as a result of the evolution of STELLA-mediated suppression of the maintenance methylation machinery. In support of this idea, ectopic expression of STELLA in medaka embryos, resulted in drastic demethylation (Mulholland et al., 2020). Intriguingly, Afrotherian and Xenarthan mammals also appear to lack a *STELLA* orthologue (Ensembl release 99, Cunningham et al., 2019). It is therefore interesting to speculate whether these the early embryos of these species would undergo a GWR event.

6.1.2.2 Ultimate causes

It is interesting to consider what function GWR may have evolved to serve in eutherian early embryos. Global removal of methylation is clearly not an absolute prerequisite for the establishment of totipotency, because in zebrafish and opossums, embryonic development proceeds without a sustained phase of hypomethylation. It has been suggested that the rapid loss of methylation in mammals enables EGA to occur earlier than in non-mammals, thereby facilitating the lineage-specification event required for the formation of the extra-embryonic tissues (Seisenberger et al., 2013, Hackett and Surani, 2013). However, the findings of this thesis demonstrate that in opossums EGA occurs in the context of a highly methylated genome (Figure 55a), meaning that global demethylation is not a prerequisite for EGA in mammals, nor for the eventual formation of extra-embryonic lineages.

One potential explanation for the difference between eutherian and metatherian mammals could be the disparity in implantation timing between these two groups. In eutherians, implantation occurs concomitant with gastrulation, whereas in metatherians implantation occurs later, once much of organogenesis has already been completed. It is conceivable that eutherian mammals must more rapidly establish extra-embryonic lineages competent for implantation, while in metatherians, the extra-embryonic tissues can afford to develop more slowly.

How would loss of methylation contribute to readying the embryo for implantation? During early eutherian development, there is widespread transcription of transposable elements (Rodriguez-Terrones and Torres-Padilla, 2018). Though some of these loci

merely represent instances of the inherent enhancer-like features of transposable elements (Todd et al., 2019), in many other instances, these features have been co-opted as regulatory elements. For example, the oocyte transcriptome contains many transcripts originating from LTRs (Veselovska et al., 2015), the two-cell stage (2C) transcriptome comprises many transcripts utilising ERVs as regulatory elements (Macfarlan et al., 2012, Eckersley-Maslin et al., 2019), and ERVs have been co-opted as enhancer elements in the placenta (Chuong et al., 2013). In eutherian mammals, the extra-embryonic tissues are hypomethylated relative to the embryonic lineage (Monk et al., 1987). It has been shown that embryonic stem cells lacking DNA methylation are competent to differentiate into trophectoderm derivatives, suggesting that the hypomethylated state is functionally associated with the formation of the extra-embryonic lineages (Ng et al., 2008). A correlation was noted between tissues that display hypomethylation and those in which novel, species-specific transposons act as regulatory elements (Chuong et al., 2013). In addition, the activation of the 2C transcriptional state is linked to hypomethylation (Eckersley-Maslin et al., 2016). Notably, the methylation landscape in oocytes is driven by LTR-derived transcription, and the methylation pattern at many of these loci is retained in the mouse blastocyst, and in the human placenta (Brind'Amour et al., 2018). Retention of oocyte-derived methylation patterns at some loci was shown to be necessary for correct gene expression in the development of the trophoblast (Branco et al., 2016). Therefore, the permissive chromatin state in hypomethylated cells releases transposable elements from repression and provides material for the evolution of novel transcriptional programmes important for the development of early embryos and the placenta (Rodriguez-Terrones and Torres-Padilla, 2018).

It is currently unknown whether the early metatherian embryo enacts a similar transposon-regulated transcriptional programme. It is notable that a decrease in global methylation level is observed in the opossum embryo around the time of lineage specification. Hypomethylation of the extra-embryonic lineages was reported to occur in gastrulation stage opossum embryos (Stevens et al., 1988). The overall decrease in methylation observed in opossum embryos from E5.5-7.5 could therefore represent a loss of methylation in the extra-embryonic lineages that are forming at this time (Figure 55b). In an alternative scenario, both the embryonic and extra-embryonic cells could undergo

demethylation, followed by remethylation in the embryonic lineage (Figure 55c). Future work analysing the methylation state of opossum embryonic and extra-embryonic lineages separately will resolve these possible scenarios. Integration of the recently generated opossum embryo transcriptome dataset (Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished), could additionally reveal whether repetitive elements in the opossum genome are released from repression during early development. The application of low-input techniques for profiling histone modifications and chromatin accessibility could further reveal the landscape of regulatory elements associated with early development in metatherians.

These findings suggest that while a hypomethylated state might also be required for the formation of extra-embryonic lineages in metatherians, this state would seem to be established late in development, and potentially only in the extra-embryonic lineage. Considering that only a subtle loss of methylation is observed in the opossum at the time of lineage specification, it is also likely that demethylation affects only a targeted set of loci rather than the entire genome. The slower timing and targeted nature of these methylation changes relative to eutherian mammals are consistent with the idea that metatherians, thanks to their later implantation, do not resort to global demethylation in order to ready the genome for lineage specification.

6.2 X chromosome inactivation

In this thesis, I examined X-chromosome DNA methylation in the opossum *M. domestica* as a representative metatherian mammal. I also examined X-chromosome methylation in mice, which represented eutherian mammals and served as an important validation control. Initially, I profiled X-chromosome methylation using RRBS in male and female adult tissues of both species, and subsequently confirmed and extended the RRBS results using allele-specific WGBS. Integration of allele-specific RNA-seq permitted the correlation of methylation patterns to gene expression status. Finally, I surveyed the pattern of DNA methylation on the opossum X chromosome during the process of XCI in early embryogenesis.

6.2.1 Role of DNA methylation in maintaining silencing

In eutherian mammals, hypermethylation of CGIs on the Xi is necessary for the maintenance of silencing (Yen et al., 1984, Sado et al., 2000; Figure 56a). Conversely, the Xa is reported to be hypermethylated overall relative to the Xi (Hellman and Chess, 2007, Weber et al., 2005, Gendrel et al., 2012, Keown et al., 2017, Duncan et al., 2018). In my examination of methylation in mouse samples, I observed CGI hypermethylation on the Xi, and by analysing the Xi and Xa separately using SNPs I showed that the methylation of XCI-escape genes resembles the active allele (Figure 56a). With these results I validated that the methods employed in this thesis can reproduce well-reported findings.

Profiling of the methylation patterns of adult opossum tissues showed that promoters on the Xi were not subject to repressive DNA methylation, in line with previous findings (Kaslow and Migeon, 1987, Loebel and Johnston, 1996, Wang et al., 2014b, Waters et al., 2018; Figure 56b). Methylation at genic regions on the Xa resembled the typical distribution found on autosomes, with low methylation at the promoter and higher methylation over the gene body and in intergenic regions (Figure 56b). Genes that were classified as escaping XCI, either in a previous report (Wang et al., 2014b), or in this thesis, displayed a methylation distribution similar to the allele on the Xa. Waters *et al* (2018) similarly found that promoters across the Xi were hypomethylated, with the

exception of XCI-escape genes. They also observed that the regions adjacent to promoters were lowly methylated on the Xi, and suggested that this may have a regulatory function in Xi-silencing. Based on the analysis presented in this thesis, which extends beyond promoter regions, I suggest that because the DNA hypomethylation extends across the entire Xi, it is unlikely to confer a specific, promoter-related regulatory function. Rather, it seems that some aspect of the unique state adopted by the Xi is refractory to DNA methylation.

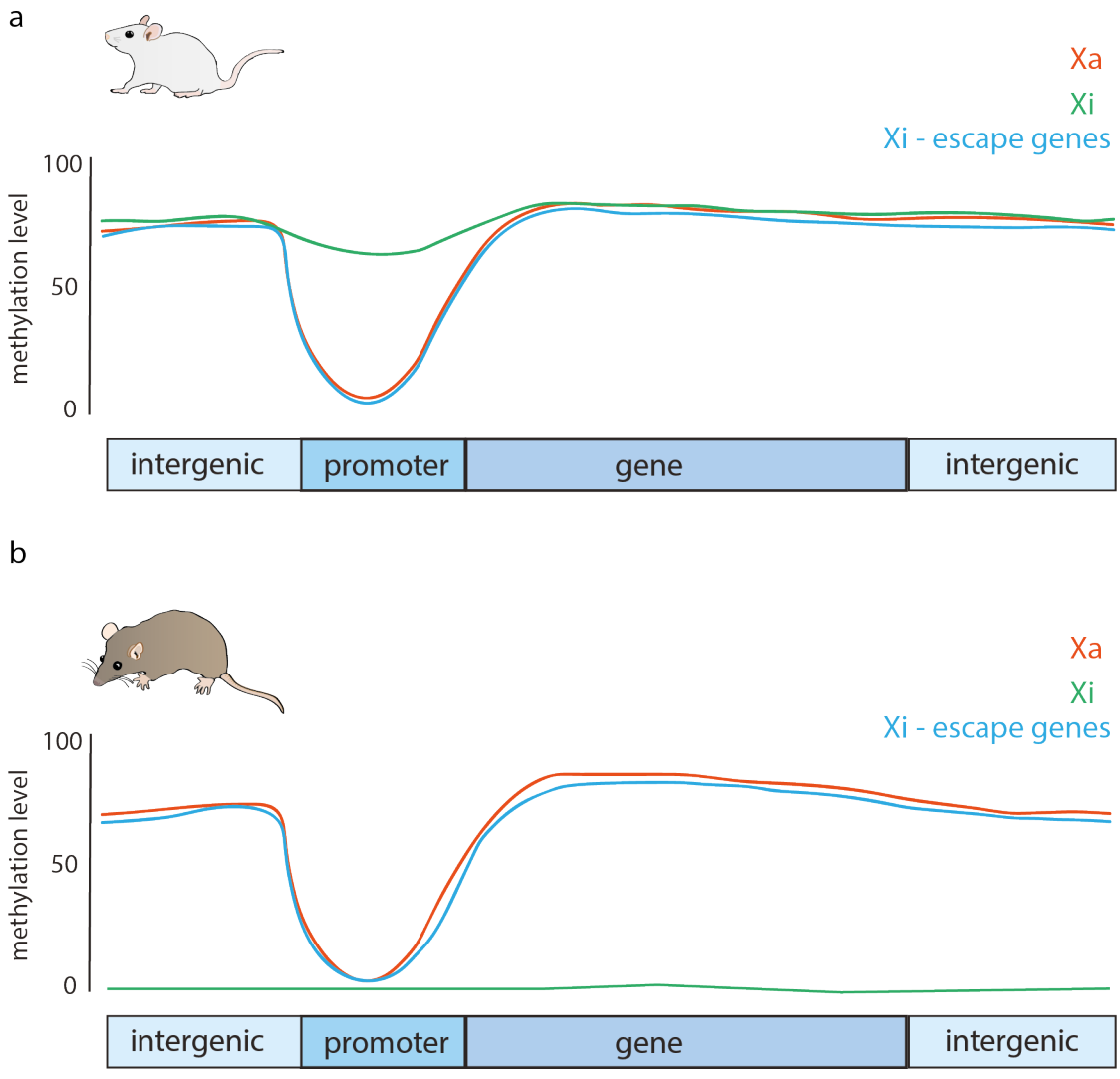


Figure 56: X-chromosome methylation landscape of therian mammals.
Schematic representing the methylation profiles of the active X (Xa) and inactive X (Xi) in mouse and opossum. **a**: Mouse. **b**: Opossum.

Analysis of DNA methylation in opossum embryos showed that the Xi is not inherited in a hypomethylated state from the paternal gamete, but rather must lose methylation during embryogenesis after XCI is initiated. The metatherian Xi replicates late in S-phase (Schneider and Rieke, 1967, Robinson et al., 1994), which suggests that maintenance methylation could be excluded from the Xi if the nuclear localisation of UHRF1 and DNMT1 is temporally regulated. It would be interesting to examine the sub-nuclear localisation of the methylation machinery in opossum embryos, to determine if the Xi is somehow excluded from maintenance methylation processes.

It was previously suggested that the absence of CGI hypermethylation on the metatherian Xi resulted in a more labile, 'leaky' silencing than in eutherian XCI, but this assumption was shown to be untrue upon examination of the expression of a wider range of X-linked genes (Mahadevaiah et al., 2009, Julien et al., 2012). Therefore, DNA methylation at promoters is not strictly necessary for robust silencing in metatherians, which differs from eutherians where loss of promoter methylation disrupts maintenance of Xi-silencing. Perhaps this reflects the fact that in eutherians, Xi genes are silent at the time of global remethylation, and are therefore subject to *de novo* methylation in the same manner as other loci across the genome. In contrast, as metatherians do not seem to undergo a global methylation reprogramming event, the opportunity to acquire Xi-hypermethylation may not have arisen during metatherian evolution.

6.2.2 **A potential mechanism for the imprinted expression of *RSX***

Metatherian XCI is imprinted such that the lncRNA *RSX* is expressed exclusively from the X_p, where it is presumed to enact chromosome-wide silencing in *cis* (Grant et al., 2012). It has recently been demonstrated that an antisense lncRNA, *XSR*, is expressed exclusively from the X_m in opossum oocytes and embryos (Shantha Mahadevaiah and Mahesh Sangrithi, Turner laboratory, unpublished). Therefore, *RSX* and *XSR* are reciprocally imprinted lncRNAs, but the molecular nature of this imprint remains to be defined. The methylation patterns in opossum gametes and embryos revealed in this thesis allows a putative mechanism of imprinting to be proposed (Figure 57).

Under this model, methylation is established at the *RSX* promoter in the oocyte due to transcription of *XSR* through the locus, while the *XSR* promoter remains unmethylated (Figure 57a). I propose that neither promoter is methylated in sperm, by inference from the methylation pattern of later embryonic stages (Figure 57b). After fertilisation, *RSX* promoter methylation is sufficient to prevent *RSX* expression from X_m in the embryo, thereby preventing X_m silencing (Figure 57c-d). Meanwhile, the unmethylated *XSR* promoter permits transcription of *XSR* from the X_m upon EGA at E3.5 (Figure 57c-d). Conversely, in female embryos the *RSX* promoter is inherited in an unmethylated state from the sperm, thereby permitting *RSX* expression from X_p at EGA, and resulting in X_p silencing in *cis* (Figure 57d).

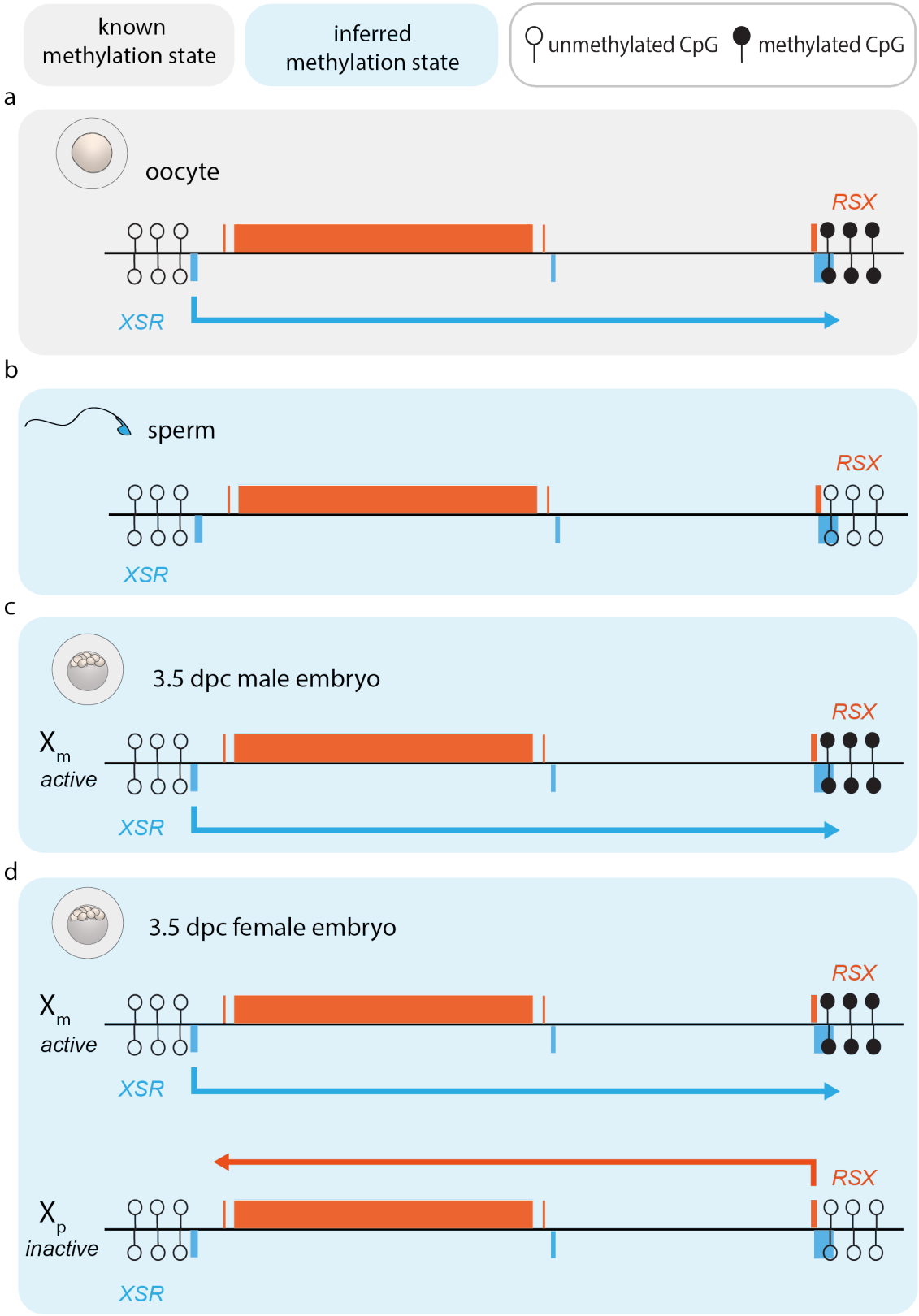

known methylation state
inferred methylation state
unmethylated CpG
methylated CpG
a
oocyte
RSX
XSR
b
sperm
RSX
XSR
c
3.5 dpc male embryo
RSX
Xm active
XSR
d
3.5 dpc female embryo
RSX
Xm active
XSR
RSX
Xp inactive
XSR

Figure 57: Potential mechanism for imprinting of the *RSX/XSR* locus. Schematic illustrating a potential methylation-dependent system of imprinting at the opossum *RSX/XSR* locus. Grey boxes indicate known methylation patterns, while blue boxes represent inferred or hypothetical methylation patterns. a: In the opossum oocyte, the *RSX* promoter is methylated, while the *XSR* promoter is unmethylated. Transcription of *XSR* may deposit methylation across the *RSX* promoter. b: By inference from methylation pattern in adults, it is likely that both the *RSX* and *XSR* promoters are unmethylated in sperm. c: Retention of oocyte-established methylation at the *RSX* promoter in male embryos could prevent *RSX* expression, while the unmethylated *XSR* promoter would permit transcription upon activation of the embryonic genome (EGA) at E3.5. d: In female embryos, the maternal X chromosome (Xm) would also retain the oocyte-established methylation, thereby expressing XSR and repressing *RSX*. Conversely, on the paternal X chromosome (Xp), the absence of methylation at the *RSX* promoter would allow *RSX* expression at EGA, resulting in *RSX* coating and X chromosome inactivation in *cis*.

This model leaves open two important questions. Firstly, what function is served by continued *XSR* expression in the embryo, given the *RSX* promoter imprint is already established in the oocyte? And secondly, if the promoter of *XSR* is unmethylated on both X_m and X_p , what prevents *XSR* expression from X_p? I suggest that continued *XSR* expression from X_m in the embryo may serve to reinforce the reciprocal repression of *RSX,* to avoid the risk of biallelic *RSX* expression. Such repression could occur in a manner analogous to regulation of imprinted loci by lncRNAs. For example, transcription across the *RSX* promoter could prevent *RSX* expression, akin to the *Igf2r/Airn* paradigm (Latos et al., 2012). In future, genetic or epigenetic perturbation of *RSX* and *XSR* in the developing opossum embryo will define the roles these lncRNAs play in metatherian XCI. The identification here of putative key regulatory regions at the *RSX* and *XSR* promoters provides a focus for such studies.

Chapter 7. Appendix

7.1 Composition of solutions used in this thesis

Tris-acetate Buffer (TAE).
40 mM Tris, 20 mM acetic acid, 1 mM EDTA.

Biggers-Whitten-Whittingham Buffer (BWW).
95mM NaCl, 44uM sodium lactate, 25mM NaHCO3, 20mM HEPES, 5.6mM D-glucose, 4.6 mM KCL, 1.7mM CaCL2, 1.2mM KH2PO4, 1.2 mM MgSO4, 0.27 mM sodium pyruvate, 0.3% (w/v) bovine serum albumin.

TYH Buffer.
119.37 mM NaCl, 4.78 mM KCl, 1.71 mM $CaCl_2$•$2H_2O$, 1.19 mM KH_2PO_4, 1.19 mM $MgSO_4$•$7H_2O$, 25.07 mM $NaHCO_3$, 1mM $C_3H_3NaO_3$, 4 g/L bovine serum albumin.

Saline-sodium-citrate Buffer (SSC).
150mM NaCl, 15mM sodium citrate.

Sperm Lysis Buffer.
2% (v/v) SDS, 20 mM DTT, 2 mM $CaCl_2$, 1600 units Proteinase K.

Polyethylene glycol Buffer (PEG).
18% (weight/volume) PEG, 1 mM EDTA, 10 mM Tris-HCl, 2.5 M NaCl, 0.05% (volume/volume) Tween-20. 0.2 micron filtered and UV-irradiated for 1 hour.

7.2 RT-PCR primers used in this thesis

Presented here are the sequences of oligonucleotide primers used to prolife the expression patterns of opossum of Y-linked genes (Chapter 3)

Table 11: RT-PCR primers used to profile expression of Y-linked genes.

Genes	Exon	Primer sequence	Annealing temperature (°C)
ATRY	exon 1	GAAAAGCAGCATGGGCAGAAT	67
	exon 5	TTCTGGCATCACAGGGTAGTG	
FAM122AY	exon 7	AGAAGCCATAGTCCCATCAACT	67
	exon 5	ACTTTAGCTGGTGAACTGCCA	
HCFC1Y	exon 15	CAAGTGGGCACCTCAACACT	69
	exon 16	TGCCATTCCTGGGTCCAATG	
HMGB3Y	exon 2	TAATGCTCCCAAACGCCCTC	66
	exon 4	CCTTTGGACTTGTAATCAGCAACA	
HSFY	exon 1	TGTGTGATGTCGGCTCAGTAG	68
	exon 2	TACCCACCCTTCTCTTCATCCT	
HUWE1	exon 16	AGCAGCATAGAGCCACCAAG	68
	exon 17	CAGCCAAAAAGGAGGGGTCT	

KDM5DY	exon 14	AGAGGAACGACAGCTACGGA	68
	exon 15	TGTTTGCTTCGAGGGCACTT	
MECP2Y	exon 1	TGGTAGCTGTGATGTTAGGGC	68
	exon 2	AACGCCTCTGTTTGGGAGAG	
OTUD5Y	exon 2	GGAGGATGGGGCTTGCTTAT	68
	exon 4	CATGGCCTGCATCTCAATGTG	
PHF6Y	exon 2	CGGTCACCCCACAGAAGTAA	68
	exon 4	ATAGTTGCACCAGGTTGGCT	
RBM10Y	exon 16	GCCCACAGTCAGCAATACCT	68
	exon 17	CACGACGCTCATCCTCCTTT	
RBMY	exon 6	CCCCACCTCCTAGGGACTAC	68
	exon 8	CTCTTGGTGTTCCTCGGCTT	
RPL10Y	exon 3	TTTGGAGGCTGCACGAATCT	67
	exon 4	GCGCAAAGCCTCAATCACAT	
TFE3Y	exon 2	GCAGAAAGAGCAACAACGGG	68
	exon 3	GGGATCCCCAGGGTAGGATT	
THOC2Y	exon 18	TGCGGCGAAGATGGAAGAAA	68
	exon 20	CCAGTCGAAGACCTGGTGAC	
UBE1Y1	exon 18	TGACTGGACTGACTGTGTGC	68
	exon 20	GGGCTGGGCAACATAGTTCT	
SRY	exon 1	TGTGGTCAAGGAGTCAACGG	68
	exon 1	TACCGAAGTGCGTGGTACAG	
TUBULIN GAMMA-1 chain-like	exon 11	GAAGTCGCCCTATCTGCCAT	68
	exon 10	GGTAGGCTATCACAGAGGCAC	

7.3 Lists of putative XCI escape genes in opossum

Presented here are lists of genes characterised as putative XCI escape genes based on their female:male methylation ratios (f:m methylation ratio > 0.8, Chapter 3).

Table 12: List of putative XCI escape genes in opossum brain.

Gene name	Gene ID	Female:male methylation ratio	Escape status (Wang *et al* 2014)
-	ENSMODG00000008271	0.920540091	non-escaper/unknown
BCAP31	ENSMODG00000008300	0.857869831	non-escaper/unknown
SSR4	ENSMODG00000008581	1.087962963	non-escaper/unknown
-	ENSMODG00000002278	0.962015267	escaper
NAA10	ENSMODG00000011547	1.094996526	non-escaper/unknown
RENBP	ENSMODG00000011560	0.837369974	non-escaper/unknown
HCFC1	ENSMODG00000011597	1.155288012	escaper
IRAK1	ENSMODG00000011627	1.072465202	escaper
MECP2	ENSMODG00000011643	1.126913656	escaper
-	ENSMODG00000034805	1.217789275	non-escaper/unknown

THOC2	ENSMODG00000014984	0.985297177	non-escaper/unknown
-	ENSMODG00000030670	0.866757866	non-escaper/unknown
-	ENSMODG00000020843	0.885258391	non-escaper/unknown
-	ENSMODG00000033237	0.942748092	non-escaper/unknown
-	ENSMODG00000008710	1.072491995	escaper
-	ENSMODG00000013238	1.062486353	non-escaper/unknown
-	ENSMODG00000008752	1.052270358	non-escaper/unknown
DNASE1L1	ENSMODG00000008763	0.803537674	non-escaper/unknown
-	ENSMODG00000008851	0.885773842	non-escaper/unknown
SLC10A3	ENSMODG00000009006	0.866991272	non-escaper/unknown
GDI1	ENSMODG00000008798	0.91333337	non-escaper/unknown
PLXNA3	ENSMODG00000008980	1.009917042	escaper
UBL4A	ENSMODG00000008997	1.10477194	non-escaper/unknown
G6PD	ENSMODG00000009154	0.933834275	escaper
IKBKG	ENSMODG00000009175	1.030082328	escaper
-	ENSMODG00000009181	0.956199658	non-escaper/unknown
-	ENSMODG00000025849	1.085714286	non-escaper/unknown
PHF6	ENSMODG00000015259	1.436256219	escaper
-	ENSMODG00000015247	0.853124529	non-escaper/unknown
FAM122B	ENSMODG00000015195	1.259194742	escaper
-	ENSMODG00000035414	1.095156944	non-escaper/unknown
-	ENSMODG00000015185	1.02912	escaper
-	ENSMODG00000025280	1.455749868	non-escaper/unknown
-	ENSMODG00000005620	0.851813299	non-escaper/unknown
DKC1	ENSMODG00000009243	1.256766297	escaper
MTMR1	ENSMODG00000005714	0.920546173	escaper
-	ENSMODG00000028278	0.913857391	escaper
-	ENSMODG00000032212	0.93955176	non-escaper/unknown
-	ENSMODG00000036135	0.897492165	non-escaper/unknown
-	ENSMODG00000005932	1.136363636	non-escaper/unknown
-	ENSMODG00000027698	1.111111111	non-escaper/unknown
-	ENSMODG00000031159	1.025	non-escaper/unknown
-	ENSMODG00000004295	0.825613306	non-escaper/unknown
ATRX	ENSMODG00000003920	0.979571512	escaper
MAGT1	ENSMODG00000003944	0.81524966	non-escaper/unknown
-	ENSMODG00000005175	0.866807611	non-escaper/unknown
-	ENSMODG00000029586	1.011969951	non-escaper/unknown
-	ENSMODG00000032286	0.862779277	non-escaper/unknown
OGT	ENSMODG00000004615	1.182574073	non-escaper/unknown
-	ENSMODG00000004570	0.974256706	escaper

ITGB1BP2	ENSMODG00000004464	0.992058212	non-escaper/unknown
NONO	ENSMODG00000004445	1.052939245	escaper
-	ENSMODG00000005121	0.905106075	non-escaper/unknown
-	ENSMODG00000004864	0.844004026	non-escaper/unknown
P2RY4	ENSMODG00000004749	0.827067669	non-escaper/unknown
-	ENSMODG00000004701	1.073964342	non-escaper/unknown
MON1B	ENSMODG00000010558	0.811323162	non-escaper/unknown
-	ENSMODG00000010111	0.907490641	non-escaper/unknown
ARAF	ENSMODG00000010855	1.944444444	non-escaper/unknown
CDK16	ENSMODG00000010767	0.803972559	non-escaper/unknown
USP11	ENSMODG00000010726	0.842741935	non-escaper/unknown
UBA1	ENSMODG00000010677	1.223662514	non-escaper/unknown
RBM10	ENSMODG00000010640	1.084028765	non-escaper/unknown
IMP4	ENSMODG00000010577	0.816208452	non-escaper/unknown
CXXC1	ENSMODG00000010555	1.064758765	non-escaper/unknown
-	ENSMODG00000010534	1.046791875	non-escaper/unknown
GRM6	ENSMODG00000010422	1.147802158	non-escaper/unknown
SUV39H1	ENSMODG00000010324	0.944120354	non-escaper/unknown
WAS	ENSMODG00000010317	0.81411686	non-escaper/unknown
TBC1D25	ENSMODG00000010259	0.896499738	non-escaper/unknown
HDAC6	ENSMODG00000010166	0.83493115	non-escaper/unknown
-	ENSMODG00000010063	1.197916667	non-escaper/unknown
PIM2	ENSMODG00000010037	1.137062101	non-escaper/unknown
-	ENSMODG00000031886	1.752475248	non-escaper/unknown
KCND1	ENSMODG00000010033	1.393137255	non-escaper/unknown
GRIPAP1	ENSMODG00000010018	0.892212883	non-escaper/unknown
CCDC120	ENSMODG00000009910	1.573172949	non-escaper/unknown
CACNA1F	ENSMODG00000009894	0.90734231	non-escaper/unknown
CCDC22	ENSMODG00000009860	1.042927159	non-escaper/unknown
-	ENSMODG00000009847	0.96803503	non-escaper/unknown
-	ENSMODG00000027767	0.856388474	non-escaper/unknown
-	ENSMODG00000009765	1.109247516	non-escaper/unknown
-	ENSMODG00000009707	1.094604493	non-escaper/unknown
RRAGB	ENSMODG00000011337	0.839813101	non-escaper/unknown
YIPF6	ENSMODG00000011692	0.958536585	escaper
-	ENSMODG00000011663	1.525624568	non-escaper/unknown
-	ENSMODG00000011746	0.947743601	escaper
-	ENSMODG00000011785	1.162392833	non-escaper/unknown
-	ENSMODG00000011798	0.825021133	non-escaper/unknown

Table 13: List of putative XCI escape genes in opossum liver.

Gene name	Gene ID	Female:male methylation ratios	Escape status (Wang *et al* 2014)
-	ENSMODG00000007462	0.8896978	non-escaper/unknown
BCAP31	ENSMODG00000008300	0.88706272	non-escaper/unknown
IDH3G	ENSMODG00000008568	0.82079639	non-escaper/unknown
SSR4	ENSMODG00000008581	0.95552297	non-escaper/unknown
-	ENSMODG00000002278	0.96931474	escaper
-	ENSMODG00000021344	0.93246618	non-escaper/unknown
NAA10	ENSMODG00000011547	1.12815849	non-escaper/unknown
RENBP	ENSMODG00000011560	0.93418584	non-escaper/unknown
HCFC1	ENSMODG00000011597	1.10729264	escaper
IRAK1	ENSMODG00000011627	1.20858394	escaper
MECP2	ENSMODG00000011643	1.13315195	escaper
-	ENSMODG00000034805	0.80406684	non-escaper/unknown
-	ENSMODG00000029915	0.90648051	non-escaper/unknown
THOC2	ENSMODG00000014984	0.92529872	non-escaper/unknown
-	ENSMODG00000033350	0.96875	non-escaper/unknown
-	ENSMODG00000007188	0.93043623	non-escaper/unknown
-	ENSMODG00000017105	0.9280303	non-escaper/unknown
-	ENSMODG00000028654	1.61347518	non-escaper/unknown
-	ENSMODG00000029677	0.92544317	non-escaper/unknown
FAM199X	ENSMODG00000010141	0.90116279	non-escaper/unknown
CUL4B	ENSMODG00000010345	0.80972885	non-escaper/unknown
-	ENSMODG00000008710	0.97066301	escaper
SASH3	ENSMODG00000013221	0.81673843	non-escaper/unknown
-	ENSMODG00000013238	1.11614182	non-escaper/unknown
-	ENSMODG00000008752	1.09704782	non-escaper/unknown
SLC10A3	ENSMODG00000009006	0.86226089	non-escaper/unknown
ATP6AP1	ENSMODG00000008869	0.81366798	non-escaper/unknown
STK26	ENSMODG00000015323	1.03512481	non-escaper/unknown
FRMD7	ENSMODG00000015320	0.82132245	escaper
RAP2C	ENSMODG00000015308	1.23833509	non-escaper/unknown
PLXNA3	ENSMODG00000008980	0.88092368	escaper
G6PD	ENSMODG00000009154	1.10092204	escaper
IKBKG	ENSMODG00000009175	1.05673846	escaper
-	ENSMODG00000009181	0.89000139	non-escaper/unknown
-	ENSMODG00000025933	1.18309859	non-escaper/unknown
-	ENSMODG00000025410	0.84589485	non-escaper/unknown
PHF6	ENSMODG00000015259	1.49758069	escaper

-	ENSMODG00000015247	0.94136606	non-escaper/unknown
FAM122B	ENSMODG00000015195	0.96769054	escaper
-	ENSMODG00000015185	0.97375037	escaper
FMR1	ENSMODG00000005550	0.94881859	non-escaper/unknown
-	ENSMODG00000005597	0.83992739	non-escaper/unknown
DKC1	ENSMODG00000009243	1.33582018	escaper
MPP1	ENSMODG00000009267	0.92058559	non-escaper/unknown
MTMR1	ENSMODG00000005714	1.03397879	escaper
-	ENSMODG00000028278	1.14044596	escaper
-	ENSMODG00000032212	0.95758243	non-escaper/unknown
-	ENSMODG00000036135	0.9306645	non-escaper/unknown
HMGB3	ENSMODG00000005735	0.85433273	escaper
-	ENSMODG00000005932	1.51623282	non-escaper/unknown
-	ENSMODG00000025405	0.83189189	non-escaper/unknown
-	ENSMODG00000027698	1.17197452	non-escaper/unknown
-	ENSMODG00000004295	0.84181216	non-escaper/unknown
ATRX	ENSMODG00000003920	1.00358528	escaper
-	ENSMODG00000004273	0.88175816	non-escaper/unknown
-	ENSMODG00000011507	0.8126399	non-escaper/unknown
-	ENSMODG00000033356	0.81174676	non-escaper/unknown
-	ENSMODG00000005175	1.01058445	non-escaper/unknown
-	ENSMODG00000031809	0.88986355	non-escaper/unknown
-	ENSMODG00000030571	0.98140541	non-escaper/unknown
-	ENSMODG00000029586	0.91601137	non-escaper/unknown
-	ENSMODG00000032286	0.94148862	non-escaper/unknown
OGT	ENSMODG00000004615	1.22545024	non-escaper/unknown
-	ENSMODG00000004570	1.03822637	escaper
ITGB1BP2	ENSMODG00000004464	0.83609447	non-escaper/unknown
NONO	ENSMODG00000004445	1.01773753	escaper
-	ENSMODG00000004864	0.83035921	non-escaper/unknown
-	ENSMODG00000004701	0.933242	non-escaper/unknown
-	ENSMODG00000027596	0.92909201	non-escaper/unknown
MON1B	ENSMODG00000010558	0.86444722	non-escaper/unknown
-	ENSMODG00000010111	0.94230715	non-escaper/unknown
ARAF	ENSMODG00000010855	1.15384615	non-escaper/unknown
CFP	ENSMODG00000010789	0.81007752	non-escaper/unknown
CDK16	ENSMODG00000010767	0.81167171	non-escaper/unknown
USP11	ENSMODG00000010726	0.83333333	non-escaper/unknown
UBA1	ENSMODG00000010677	1.0464785	non-escaper/unknown
RBM10	ENSMODG00000010640	0.86973078	non-escaper/unknown

IMP4	ENSMODG00000010577	0.89461576	non-escaper/unknown
CXXC1	ENSMODG00000010555	1.12719466	non-escaper/unknown
-	ENSMODG00000010534	1.08475797	non-escaper/unknown
GRM6	ENSMODG00000010422	1.15832243	non-escaper/unknown
SUV39H1	ENSMODG00000010324	0.93801757	non-escaper/unknown
SLC38A5	ENSMODG00000010292	0.90201081	non-escaper/unknown
HDAC6	ENSMODG00000010166	0.80242807	non-escaper/unknown
-	ENSMODG00000010069	0.88079766	non-escaper/unknown
-	ENSMODG00000010063	1.41265643	non-escaper/unknown
PIM2	ENSMODG00000010037	0.91211702	non-escaper/unknown
KCND1	ENSMODG00000010033	0.96818891	non-escaper/unknown
GRIPAP1	ENSMODG00000010018	0.98470736	non-escaper/unknown
TFE3	ENSMODG00000009925	1.24820531	non-escaper/unknown
CCDC120	ENSMODG00000009910	0.99617492	non-escaper/unknown
CACNA1F	ENSMODG00000009894	0.89121016	non-escaper/unknown
CCDC22	ENSMODG00000009860	1.13808701	non-escaper/unknown
-	ENSMODG00000009847	0.89341398	non-escaper/unknown
-	ENSMODG00000027767	0.92566938	non-escaper/unknown
-	ENSMODG00000009765	1.17919809	non-escaper/unknown
-	ENSMODG00000009707	0.99524639	non-escaper/unknown
PIN4	ENSMODG00000010913	1.24238279	non-escaper/unknown
-	ENSMODG00000026293	0.81393167	non-escaper/unknown
RRAGB	ENSMODG00000011337	0.8257765	non-escaper/unknown
YIPF6	ENSMODG00000011692	1.01190605	escaper
WDR44	ENSMODG00000011281	0.80247189	non-escaper/unknown
-	ENSMODG00000011590	1.09619053	non-escaper/unknown
-	ENSMODG00000011691	1.26243321	non-escaper/unknown
-	ENSMODG00000011746	0.94477226	escaper
-	ENSMODG00000011785	1.55408221	non-escaper/unknown

7.4 Oligonucleotide sequences used in the preparation of BS-seq libraries

Table 14: Oligonucleotides used in preparation of BS-seq libraries.

Name	Sequence	Supplier
Preamp Primer	CTACACGACGCTCTTCCGATCTNNNNNN	IDT
Adapter 2 Oligo for NEB indices	CAGACGTGTGCTCTTCCGATCTNNNNNN	IDT

7.5 Genes identified as XCI escapers by allele-specific RNA-seq.

Presented here are lists of genes identified as escaping XCI based on the analysis of allele-specific RNA-seq (allelic ratio > 0.1, Chapter 4).

Table 15: Mouse XCI escape genes.

Gene name	Gene ID	Brain allelic ratio	Liver allelic ratio	Previously described as an escape gene
Cfp	ENSMUSG00000001128	-	0.184782609	Berletch *et al* 2015
Ddx3x	ENSMUSG00000000787	0.261765198	0.366187543	Berletch *et al* 2015
Eif2s3x	ENSMUSG00000035150	0.281459419	0.282566429	Berletch *et al* 2015
Ftx	ENSMUSG00000086370	-	0.268041237	Berletch *et al* 2015
Jpx	ENSMUSG00000097571	0.103896104	0.181818182	Tian *et al* 2010
Kdm5c	ENSMUSG00000025332	0.203912716	0.2760181	Berletch *et al* 2015
Kdm6a	ENSMUSG00000037369	0.265017668	0.35890411	Berletch *et al* 2015
Lpar4	ENSMUSG00000049929	0.101449275	-	Wang *et al* 2017
Pbdc1	ENSMUSG00000031226	0.192105263	0.121568627	Berletch *et al* 2015
Tmem29	ENSMUSG00000041353	-	0.116363636	Berletch *et al* 2015
Xist	ENSMUSG00000086503	0.785707525	0.922842197	Berletch *et al* 2015
5530601H04Rik	ENSMUSG00000087174	0.221216041	0.311827957	Berletch *et al* 2015

Table 16: Opossum XCI escape genes.

Gene name	Gene ID	Brain allelic ratio	Liver allelic ratio	Spleen allelic ratio	Previously identified as an escape gene
RBMX	ENSMODG00000014054	0.19	0.3	0.33	Wang *et al* 2014
G-protein coupled receptor 12-like	ENSMODG00000022849	0.63	-	-	-
ZBTB33	ENSMODG00000025434	0.12	-	-	-
FLNA	ENSMODG00000008710	-	-	0.33	Wang *et al* 2014
RRAGB	ENSMODG00000011337	-	-	0.36	-
ARHGAP4	ENSMODG00000011526	-	-	0.14	-

7.6 Software

Presented in this appendix are version details for software used in this thesis.

Table 17: Details of software used in this thesis.

Software name	Version
FastQC	0.11.5
TrimGalore!	0.6.0
SAMTools	1.4
SNPsplit	0.3.4
Fiji/ImageJ	2.0.0
Bismark	0.18.0
HISAT2	2.1.0
Seqmonk	0.34.1
R	3.4.3
R packages	
ggplot2	3.2.0
tidyverse	1.2.1
methylKit	1.4.1
genomicRanges	1.30.3
IRanges	2.12.0
DESeq2	1.18.1
rtracklayer	1.38.3
Gviz	1.22.3
refGenome	1.7.3
Rsubread	1.28.1
ggfortify	0.4.1
ComplexHeatmap	2.5.1

Reference List

ADLER, D. A., BRESSLER, S. L., CHAPMAN, V. M., PAGE, D. C. & DISTECHE, C. M. 1991. Inactivation of the Zfx gene on the mouse X chromosome. *Proc Natl Acad Sci U S A,* 88**,** 4592-5.

ADLER, D. A., RUGARLI, E. I., LINGENFELTER, P. A., TSUCHIYA, K., POSLINSKI, D., LIGGITT, H. D., CHAPMAN, V. M., ELLIOTT, R. W., BALLABIO, A. & DISTECHE, C. M. 1997. Evidence of evolutionary up-regulation of the single active X chromosome in mammals based on Clc4 expression levels in Mus spretus and Mus musculus. *Proc Natl Acad Sci U S A,* 94**,** 9244-8.

AGULNIK, A. I., MITCHELL, M. J., MATTEI, M. G., BORSANI, G., AVNER, P. A., LERNER, J. L. & BISHOP, C. E. 1994. A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum Mol Genet,* 3**,** 879-84.

AKALIN, A., KORMAKSSON, M., LI, S., GARRETT-BAKELMAN, F. E., FIGUEROA, M. E., MELNICK, A. & MASON, C. E. 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol,* 13**,** R87.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

AMOUROUX, R., NASHUN, B., SHIRANE, K., NAKAGAWA, S., HILL, P. W., D'SOUZA, Z., NAKAYAMA, M., MATSUDA, M., TURP, A., NDJETEHE, E., ENCHEVA, V., KUDO, N. R., KOSEKI, H., SASAKI, H. & HAJKOVA, P. 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol,* 18**,** 225-233.

ANDREWS, S. 2007. *Seqmonk. A tool to visualise and analyse high throughput mapped sequence data* [Online]. Available: https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/ [Accessed].

ANDREWS, S. 2010. *FastQC: A quality control tool for high throughput sequence data.* [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed].

ARAVIN, A. A., SACHIDANANDAM, R., BOURC'HIS, D., SCHAEFER, C., PEZIC, D., TOTH, K. F., BESTOR, T. & HANNON, G. J. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell,* 31**,** 785-99.

ARNOLD, R. & SHOREY, C. D. 1985a. Structure of the oviducal epithelium of the brush-tailed possum (Trichosurus vulpecula). *J Reprod Fertil,* 73**,** 9-19.

ARNOLD, R. & SHOREY, C. D. 1985b. Structure of the uterine luminal epithelium of the brush-tailed possum (Trichosurus vulpecula). *J Reprod Fertil,* 74**,** 565-73.

AUCLAIR, G., GUIBERT, S., BENDER, A. & WEBER, M. 2014. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol,* 15**,** 545.

BAGGOTT, L. M., DAVIS-BUTLER, S. & MOORE, H. D. 1987. Characterization of oestrus and timed collection of oocytes in the grey short-tailed opossum, Monodelphis domestica. *J Reprod Fertil,* 79**,** 105-14.

BALHORN, R. 1982. A model for the structure of chromatin in mammalian sperm. *J Cell Biol,* 93**,** 298-305.

BALHORN, R. 2007. The protamine family of sperm nuclear proteins. *Genome Biol,* 8**,** 227.

BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.

BANNISTER, A. J., SCHNEIDER, R., MYERS, F. A., THORNE, A. W., CRANE-ROBINSON, C. & KOUZARIDES, T. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem,* 280**,** 17732-6.

BARAU, J., TEISSANDIER, A., ZAMUDIO, N., ROY, S., NALESSO, V., HERAULT, Y., GUILLOU, F. & BOURC'HIS, D. 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science,* 354**,** 909-912.

BARLOW, D. P., STOGER, R., HERRMANN, B. G., SAITO, K. & SCHWEIFER, N. 1991. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature,* 349**,** 84-7.

BARNETT, K. R., DECATO, B. E., SCOTT, T. J., HANSEN, T. J., CHEN, B., ATTALLA, J., SMITH, A. D. & HODGES, E. 2020. ATAC-Me Captures Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions. *Mol Cell*.

BARTOLOMEI, M. S., WEBBER, A. L., BRUNKOW, M. E. & TILGHMAN, S. M. 1993. Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev,* 7**,** 1663-73.

BARTOLOMEI, M. S., ZEMEL, S. & TILGHMAN, S. M. 1991. Parental imprinting of the mouse H19 gene. *Nature,* 351**,** 153-5.

BARTON, S. C., ARNEY, K. L., SHI, W., NIVELEAU, A., FUNDELE, R., SURANI, M. A. & HAAF, T. 2001. Genome-wide methylation patterns in normal and uniparental early mouse embryos. *Hum Mol Genet,* 10**,** 2983-7.

BARTON, S. C., SURANI, M. A. & NORRIS, M. L. 1984. Role of paternal and maternal genomes in mouse development. *Nature,* 311**,** 374-6.

BAUBEC, T., COLOMBO, D. F., WIRBELAUER, C., SCHMIDT, J., BURGER, L., KREBS, A. R., AKALIN, A. & SCHUBELER, D. 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature,* 520**,** 243-7.

BAUBEC, T., IVANEK, R., LIENERT, F. & SCHUBELER, D. 2013. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell,* 153**,** 480-92.

BEAUJEAN, N., HARTSHORNE, G., CAVILLA, J., TAYLOR, J., GARDNER, J., WILMUT, I., MEEHAN, R. & YOUNG, L. 2004. Non-conservation of mammalian preimplantation methylation dynamics. *Curr Biol,* 14**,** R266-7.

BELLOTT, D. W., HUGHES, J. F., SKALETSKY, H., BROWN, L. G., PYNTIKOVA, T., CHO, T. J., KOUTSEVA, N., ZAGHLUL, S., GRAVES, T., ROCK, S., KREMITZKI, C., FULTON, R. S., DUGAN, S., DING, Y., MORTON, D., KHAN, Z., LEWIS, L., BUHAY, C., WANG, Q., WATT, J., HOLDER, M., LEE, S., NAZARETH, L., ALFOLDI, J., ROZEN, S., MUZNY, D. M., WARREN, W. C., GIBBS, R. A., WILSON, R. K. & PAGE, D. C. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature,* 508**,** 494-9.

BEN-HATTAR, J. & JIRICNY, J. 1988. Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. *Gene,* 65**,** 219-27.

BERGALLO, H. G. & CERQUEIRA, R. 1994. Reproduction and Growth of the Opossum Monodelphis-Domestica (Mammalia, Didelphidae) in Northeastern Brazil. *Journal of Zoology,* 232**,** 551-563.

BERLETCH, J. B., MA, W., YANG, F., SHENDURE, J., NOBLE, W. S., DISTECHE, C. M. & DENG, X. 2015. Escape from X inactivation varies in mouse tissues. *PLoS Genet,* 11**,** e1005079.

BERNSTEIN, B. E., MIKKELSEN, T. S., XIE, X., KAMAL, M., HUEBERT, D. J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K., JAENISCH, R., WAGSCHAL, A., FEIL, R., SCHREIBER, S. L. & LANDER, E. S. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell,* 125**,** 315-26.

BESTOR, T. H., EDWARDS, J. R. & BOULARD, M. 2015. Notes on the role of dynamic DNA methylation in mammalian development. *Proc Natl Acad Sci U S A,* 112**,** 6796-9.

BEWICK, A. J., VOGEL, K. J., MOORE, A. J. & SCHMITZ, R. J. 2017. Evolution of DNA Methylation across Insects. *Mol Biol Evol,* 34**,** 654-665.

BIAN, C. & YU, X. 2014. PGC7 suppresses TET3 for protecting DNA methylation. *Nucleic Acids Res,* 42**,** 2893-905.

BININDA-EMONDS, O. R., CARDILLO, M., JONES, K. E., MACPHEE, R. D., BECK, R. M., GRENYER, R., PRICE, S. A., VOS, R. A., GITTLEMAN, J. L. & PURVIS, A. 2007. The delayed rise of present-day mammals. *Nature,* 446**,** 507-12.

BIRD, A., TAGGART, M., FROMMER, M., MILLER, O. J. & MACLEOD, D. 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell,* 40**,** 91-9.

BIRD, A., TAGGART, M. & MACLEOD, D. 1981. Loss of rDNA methylation accompanies the onset of ribosomal gene activity in early development of X. laevis. *Cell,* 26**,** 381-90.

BIRD, A. P. 1978. Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *J Mol Biol,* 118**,** 49-60.

BIRD, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res,* 8**,** 1499-504.

BIRD, A. P. 1986. CpG-rich islands and the function of DNA methylation. *Nature,* 321**,** 209-13.

BLEWITT, M. E., GENDREL, A. V., PANG, Z., SPARROW, D. B., WHITELAW, N., CRAIG, J. M., APEDAILE, A., HILTON, D. J., DUNWOODIE, S. L., BROCKDORFF, N., KAY, G. F. & WHITELAW, E. 2008. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet,* 40**,** 663-9.

BOCK, C., BEERMAN, I., LIEN, W. H., SMITH, Z. D., GU, H., BOYLE, P., GNIRKE, A., FUCHS, E., ROSSI, D. J. & MEISSNER, A. 2012. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell,* 47**,** 633-47.

BOGGS, B. A., CHEUNG, P., HEARD, E., SPECTOR, D. L., CHINAULT, A. C. & ALLIS, C. D. 2002. Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat Genet,* 30**,** 73-6.

BOLZER, A., KRETH, G., SOLOVEI, I., KOEHLER, D., SARACOGLU, K., FAUTH, C., MULLER, S., EILS, R., CREMER, C., SPEICHER, M. R. & CREMER, T. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol,* 3**,** e157.

BORENSZTEIN, M., OKAMOTO, I., SYX, L., GUILBAUD, G., PICARD, C., ANCELIN, K., GALUPA, R., DIABANGOUAYA, P., SERVANT, N., BARILLOT, E., SURANI, A., SAITOU, M., CHEN, C. J., ANASTASSIADIS, K. & HEARD, E. 2017. Contribution of epigenetic landscapes and transcription factors to X-chromosome reactivation in the inner cell mass. *Nat Commun,* 8**,** 1297.

BORGEL, J., GUIBERT, S., LI, Y., CHIBA, H., SCHUBELER, D., SASAKI, H., FORNE, T. & WEBER, M. 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet,* 42**,** 1093-100.

BORSANI, G., TONLORENZI, R., SIMMLER, M. C., DANDOLO, L., ARNAUD, D., CAPRA, V., GROMPE, M., PIZZUTI, A., MUZNY, D., LAWRENCE, C., WILLARD, H. F., AVNER, P. & BALLABIO, A. 1991. Characterization of a murine gene expressed from the inactive X chromosome. *Nature,* 351**,** 325-9.

BORTVIN, A., GOODHEART, M., LIAO, M. & PAGE, D. C. 2004. Dppa3 / Pgc7 / stella is a maternal factor and is not required for germ cell specification in mice. *BMC Dev Biol,* 4**,** 2.

BOSTICK, M., KIM, J. K., ESTEVE, P. O., CLARK, A., PRADHAN, S. & JACOBSEN, S. E. 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science,* 317**,** 1760-4.

BOURC'HIS, D. & BESTOR, T. H. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature,* 431**,** 96-9.

BOURC'HIS, D., XU, G. L., LIN, C. S., BOLLMAN, B. & BESTOR, T. H. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science,* 294**,** 2536-9.

BOYD, Y. & FRASER, N. J. 1990. Methylation patterns at the hypervariable X-chromosome locus DXS255 (M27 beta): correlation with X-inactivation status. *Genomics,* 7**,** 182-7.

BOYLE, P., CLEMENT, K., GU, H., SMITH, Z. D., ZILLER, M., FOSTEL, J. L., HOLMES, L., MELDRIM, J., KELLEY, F., GNIRKE, A. & MEISSNER, A. 2012. Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol,* 13**,** R92.

BRANCO, M. R., KING, M., PEREZ-GARCIA, V., BOGUTZ, A. B., CALEY, M., FINEBERG, E., LEFEBVRE, L., COOK, S. J., DEAN, W., HEMBERGER, M. & REIK, W. 2016. Maternal DNA Methylation Regulates Early Trophoblast Development. *Dev Cell,* 36**,** 152-63.

BRAUDE, P., BOLTON, V. & MOORE, S. 1988. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature,* 332**,** 459-61.

BRIND'AMOUR, J., KOBAYASHI, H., RICHARD ALBERT, J., SHIRANE, K., SAKASHITA, A., KAMIO, A., BOGUTZ, A., KOIKE, T., KARIMI, M. M., LEFEBVRE, L., KONO, T. & LORINCZ, M. C. 2018. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat Commun,* 9**,** 3331.

BROCKDORFF, N., ASHWORTH, A., KAY, G. F., COOPER, P., SMITH, S., MCCABE, V. M., NORRIS, D. P., PENNY, G. D., PATEL, D. & RASTAN, S. 1991. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature,* 351**,** 329-31.

BROWN, C. J., BALLABIO, A., RUPERT, J. L., LAFRENIERE, R. G., GROMPE, M., TONLORENZI, R. & WILLARD, H. F. 1991a. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature,* 349**,** 38-44.

BROWN, C. J., LAFRENIERE, R. G., POWERS, V. E., SEBASTIO, G., BALLABIO, A., PETTIGREW, A. L., LEDBETTER, D. H., LEVY, E., CRAIG, I. W. & WILLARD, H. F. 1991b. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature,* 349**,** 82-4.

BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol,* 109**,** 21 29 1-21 29 9.

BURGOYNE, P. S. 1982. Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Hum Genet,* 61**,** 85-90.

CAPEL, B. 2017. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat Rev Genet,* 18**,** 675-689.

CARREL, L., COTTLE, A. A., GOGLIN, K. C. & WILLARD, H. F. 1999. A first-generation X-inactivation profile of the human X chromosome. *Proc Natl Acad Sci U S A,* 96**,** 14440-4.

CARREL, L. & WILLARD, H. F. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature,* 434**,** 400-4.

CATTANACH, B. M. & KIRK, M. 1985. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature,* 315**,** 496-8.

CHAILLET, J. R., VOGT, T. F., BEIER, D. R. & LEDER, P. 1991. Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. *Cell,* 66**,** 77-83.

CHAMPROUX, A., TORRES-CARREIRA, J., GHARAGOZLOO, P., DREVET, J. R. & KOCER, A. 2016. Mammalian sperm nuclear organization: resiliencies and vulnerabilities. *Basic Clin Androl,* 26**,** 17.

CHARLESWORTH, B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol,* 6**,** 149-62.

CHAUMEIL, J., LE BACCON, P., WUTZ, A. & HEARD, E. 2006. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev,* 20**,** 2223-37.

CHEN, T., ZHANG, Y. L., JIANG, Y., LIU, S. Z., SCHATTEN, H., CHEN, D. Y. & SUN, Q. Y. 2004. The DNA methylation events in normal and cloned rabbit embryos. *FEBS Lett,* 578**,** 69-72.

CHEN, Z. & ZHANG, Y. 2019. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat Genet,* 51**,** 947-951.

CHIBA, H., HIRASAWA, R., KANEDA, M., AMAKAWA, Y., LI, E., SADO, T. & SASAKI, H. 2008. De novo DNA methylation independent establishment of maternal imprint on X chromosome in mouse oocytes. *Genesis,* 46**,** spc one.

CHOTALIA, M., SMALLWOOD, S. A., RUF, N., DAWSON, C., LUCIFERO, D., FRONTERA, M., JAMES, K., DEAN, W. & KELSEY, G. 2009. Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev,* 23**,** 105-17.

CHUONG, E. B., RUMI, M. A., SOARES, M. J. & BAKER, J. C. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet,* 45**,** 325-9.

CIRIO, M. C., MARTEL, J., MANN, M., TOPPINGS, M., BARTOLOMEI, M., TRASLER, J. & CHAILLET, J. R. 2008a. DNA methyltransferase 1o functions during preimplantation development to preclude a profound level of epigenetic variation. *Dev Biol,* 324**,** 139-50.

CIRIO, M. C., RATNAM, S., DING, F., REINHART, B., NAVARA, C. & CHAILLET, J. R. 2008b. Preimplantation expression of the somatic form of Dnmt1 suggests a role in the inheritance of genomic imprints. *BMC Dev Biol,* 8**,** 9.

CLARK, S. J., LEE, H. J., SMALLWOOD, S. A., KELSEY, G. & REIK, W. 2016. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol,* 17**,** 72.

CLARK, S. J., SMALLWOOD, S. A., LEE, H. J., KRUEGER, F., REIK, W. & KELSEY, G. 2017. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc,* 12**,** 534-547.

CLEMSON, C. M., MCNEIL, J. A., WILLARD, H. F. & LAWRENCE, J. B. 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol,* 132**,** 259-75.

COKUS, S. J., FENG, S., ZHANG, X., CHEN, Z., MERRIMAN, B., HAUDENSCHILD, C. D., PRADHAN, S., NELSON, S. F., PELLEGRINI, M. & JACOBSEN, S. E. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature,* 452.

COOMBES, C., ARNAUD, P., GORDON, E., DEAN, W., COAR, E. A., WILLIAMSON, C. M., FEIL, R., PETERS, J. & KELSEY, G. 2003. Epigenetic properties and identification of an imprint mark in the Nesp-Gnasxl domain of the mouse Gnas imprinted locus. *Mol Cell Biol,* 23**,** 5475-88.

COOPER, D. N., TAGGART, M. H. & BIRD, A. P. 1983. Unmethylated domains in vertebrate DNA. *Nucleic Acids Res,* 11**,** 647-58.

COOPER, D. W., JOHNSTON, P. G., WATSON, J. M. & GRAVES, J. A. M. 1993. X-inactivation in marsupials and monotremes. *Seminars in Developmental Biology,* 4**,** 117-128.

CORTEZ, D., MARIN, R., TOLEDO-FLORES, D., FROIDEVAUX, L., LIECHTI, A., WATERS, P. D., GRUTZNER, F. & KAESSMANN, H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature,* 508**,** 488-93.

COSTANZI, C. & PEHRSON, J. R. 1998. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature,* 393**,** 599-601.

COTTON, A. M., GE, B., LIGHT, N., ADOUE, V., PASTINEN, T. & BROWN, C. J. 2013. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol,* 14**,** R122.

COTTON, A. M., LAM, L., AFFLECK, J. G., WILSON, I. M., PENAHERRERA, M. S., MCFADDEN, D. E., KOBOR, M. S., LAM, W. L., ROBINSON, W. P. & BROWN, C. J. 2011. Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum Genet,* 130**,** 187-201.

COTTON, A. M., PRICE, E. M., JONES, M. J., BALATON, B. P., KOBOR, M. S. & BROWN, C. J. 2015. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet,* 24**,** 1528-39.

CREMER, T. & CREMER, M. 2010. Chromosome territories. *Cold Spring Harb Perspect Biol,* 2**,** a003889.

CSANKOVSZKI, G., NAGY, A. & JAENISCH, R. 2001. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J Cell Biol,* 153**,** 773-84.

CUMMINS, J. M. 1980. Decondensation of sperm nuclei of australian mammals: Effects of air drying and of calcium and magnesium. *Gamete Research,* 3**,** 351-367.

CUNNINGHAM, F., ACHUTHAN, P., AKANNI, W., ALLEN, J., AMODE, M. R., ARMEAN, I. M., BENNETT, R., BHAI, J., BILLIS, K., BODDU, S., CUMMINS, C., DAVIDSON, C., DODIYA, K. J., GALL, A., GIRON, C. G., GIL, L., GREGO, T., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., KAY, M., LAIRD, M. R., LAVIDAS, I., LIU, Z., LOVELAND, J. E., MARUGAN, J. C., MAUREL, T., MCMAHON, A. C., MOORE, B., MORALES, J., MUDGE, J. M., NUHN, M., OGEH, D., PARKER, A., PARTON, A., PATRICIO, M., ABDUL SALAM, A. I., SCHMITT, B. M., SCHUILENBURG, H., SHEPPARD, D., SPARROW, H., STAPLETON, E., SZUBA, M., TAYLOR, K., THREADGOLD, G., THORMANN, A., VULLO, A., WALTS, B., WINTERBOTTOM, A., ZADISSA, A., CHAKIACHVILI, M., FRANKISH, A., HUNT, S. E., KOSTADIMA, M., LANGRIDGE, N., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., AKEN, B. L., YATES, A. D., ZERBINO, D. R. & FLICEK, P. 2019. Ensembl 2019. *Nucleic Acids Res,* 47**,** D745-D751.

DAWLATY, M. M., BREILING, A., LE, T., BARRASA, M. I., RADDATZ, G., GAO, Q., POWELL, B. E., CHENG, A. W., FAULL, K. F., LYKO, F. & JAENISCH, R. 2014. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell,* 29**,** 102-11.

DAWLATY, M. M., BREILING, A., LE, T., RADDATZ, G., BARRASA, M. I., CHENG, A. W., GAO, Q., POWELL, B. E., LI, Z., XU, M., FAULL, K. F., LYKO, F. & JAENISCH, R. 2013. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev Cell,* 24**,** 310-23.

DAWLATY, M. M., GANZ, K., POWELL, B. E., HU, Y. C., MARKOULAKI, S., CHENG, A. W., GAO, Q., KIM, J., CHOI, S. W., PAGE, D. C. & JAENISCH, R. 2011. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell,* 9**,** 166-75.

DE IACO, A., PLANET, E., COLUCCIO, A., VERP, S., DUC, J. & TRONO, D. 2017. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet,* 49**,** 941-945.

DE IACO, A., VERP, S., OFFNER, S., GRUN, D. & TRONO, D. 2020. DUX is a non-essential synchronizer of zygotic genome activation. *Development,* 147.

DE SANTA, F., BAROZZI, I., MIETTON, F., GHISLETTI, S., POLLETTI, S., TUSI, B. K., MULLER, H., RAGOUSSIS, J., WEI, C. L. & NATOLI, G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol,* 8**,** e1000384.

DE WIT, E., VOS, E. S., HOLWERDA, S. J., VALDES-QUEZADA, C., VERSTEGEN, M. J., TEUNISSEN, H., SPLINTER, E., WIJCHERS, P. J., KRIJGER, P. H. & DE LAAT, W. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell,* 60**,** 676-84.

DEAN, W., SANTOS, F., STOJKOVIC, M., ZAKHARTCHENKO, V., WALTER, J., WOLF, E. & REIK, W. 2001. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc Natl Acad Sci U S A,* 98**,** 13734-8.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25.

DECHIARA, T. M., ROBERTSON, E. J. & EFSTRATIADIS, A. 1991. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell,* 64**,** 849-59.

DEICHMANN, U. 2016. Epigenetics: the origins and evolution of a fashionable topic. *Dev Biol*.

DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *Science,* 295**,** 1306-11.

DENG, X., BERLETCH, J. B., MA, W., NGUYEN, D. K., HIATT, J. B., NOBLE, W. S., SHENDURE, J. & DISTECHE, C. M. 2013. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell,* 25**,** 55-68.

DENG, X., HIATT, J. B., NGUYEN, D. K., ERCAN, S., STURGILL, D., HILLIER, L. W., SCHLESINGER, F., DAVIS, C. A., REINKE, V. J., GINGERAS, T. R., SHENDURE, J., WATERSTON, R. H., OLIVER, B., LIEB, J. D. & DISTECHE, C. M. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat Genet,* 43**,** 1179-85.

DESHMUKH, R. S., OSTRUP, O., OSTRUP, E., VEJLSTED, M., NIEMANN, H., LUCAS-HAHN, A., PETERSEN, B., LI, J., CALLESEN, H. & HYTTEL, P. 2011. DNA methylation in porcine preimplantation embryos developed in vivo and produced by in vitro fertilization, parthenogenetic activation and somatic cell nuclear transfer. *Epigenetics,* 6**,** 177-87.

DING, F., PATEL, C., RATNAM, S., MCCARREY, J. R. & CHAILLET, J. R. 2003. Conservation of Dnmt1o cytosine methyltransferase in the marsupial Monodelphis domestica. *Genesis,* 36**,** 209-13.

DISTECHE, C. M., ZACKSENHAUS, E., ADLER, D. A., BRESSLER, S. L., KEITZ, B. T. & CHAPMAN, V. M. 1992. Mapping and expression of the ubiquitin-activating enzyme E1 (Ube1) gene in the mouse. *Mamm Genome,* 3**,** 156-61.

DOBSON, A. T., RAJA, R., ABEYTA, M. J., TAYLOR, T., SHEN, S., HAQQ, C. & PERA, R. A. 2004. The unique transcriptome through day 3 of human preimplantation development. *Hum Mol Genet,* 13**,** 1461-70.

DODGE, J. E., RAMSAHOYE, B. H., WO, Z. G., OKANO, M. & LI, E. 2002. De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene,* 289**,** 41-8.

DU, J., JOHNSON, L. M., JACOBSEN, S. E. & PATEL, D. J. 2015. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol,* 16**,** 519-32.

DUAN, J. E., JIANG, Z. C., ALQAHTANI, F., MANDOIU, I., DONG, H., ZHENG, X., MARJANI, S. L., CHEN, J. & TIAN, X. C. 2019. Methylome Dynamics of Bovine Gametes and in vivo Early Embryos. *Front Genet,* 10**,** 512.

DUNCAN, C. G., GRIMM, S. A., MORGAN, D. L., BUSHEL, P. R., BENNETT, B. D., PROGRAM, N. C. S., ROBERTS, J. D., TYSON, F. L., MERRICK, B. A. & WADE, P. A. 2018. Dosage compensation and DNA methylation landscape of the X chromosome in mouse liver. *Sci Rep,* 8**,** 10138.

DURET, L., CHUREAU, C., SAMAIN, S., WEISSENBACH, J. & AVNER, P. 2006. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science,* 312**,** 1653-5.

ECKERSLEY-MASLIN, M., ALDA-CATALINAS, C., BLOTENBURG, M., KREIBICH, E., KRUEGER, C. & REIK, W. 2019. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev,* 33**,** 194-208.

ECKERSLEY-MASLIN, M. A., SVENSSON, V., KRUEGER, C., STUBBS, T. M., GIEHR, P., KRUEGER, F., MIRAGAIA, R. J., KYRIAKOPOULOS, C., BERRENS, R. V., MILAGRE, I., WALTER, J., TEICHMANN, S. A. & REIK, W. 2016. MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep,* 17**,** 179-192.

EDWARDS, C. A. & FERGUSON-SMITH, A. C. 2007. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol,* 19**,** 281-9.

EHRMANN, I. E., ELLIS, P. S., MAZEYRAT, S., DUTHIE, S., BROCKDORFF, N., MATTEI, M. G., GAVIN, M. A., AFFARA, N. A., BROWN, G. M., SIMPSON, E., MITCHELL, M. J. & SCOTT, D. M. 1998. Characterization of genes encoding translation initiation factor eIF-2gamma in mouse and human: sex chromosome localization, escape from X-inactivation and evolution. *Hum Mol Genet,* 7**,** 1725-37.

ELLISON, J., PASSAGE, M., YU, L. C., YEN, P., MOHANDAS, T. K. & SHAPIRO, L. 1992. Directed isolation of human genes that escape X inactivation. *Somat Cell Mol Genet,* 18**,** 259-68.

ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

FALCO, G., LEE, S. L., STANGHELLINI, I., BASSEY, U. C., HAMATANI, T. & KO, M. S. 2007. Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev Biol,* 307**,** 539-50.

FELDMANN, A., IVANEK, R., MURR, R., GAIDATZIS, D., BURGER, L. & SCHUBELER, D. 2013. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet,* 9**,** e1003994.

FELLOUS, A., LABED-VEYDERT, T., LOCREL, M., VOISIN, A. S., EARLEY, R. L. & SILVESTRE, F. 2018. DNA methylation in adults and during development of the self-fertilizing mangrove rivulus, Kryptolebias marmoratus. *Ecol Evol,* 8**,** 6016-6033.

FELSENFELD, G., BOYES, J., CHUNG, J., CLARK, D. & STUDITSKY, V. 1996. Chromatin structure and gene expression. *Proc Natl Acad Sci U S A,* 93**,** 9384-8.

FERGUSON-SMITH, A. C., CATTANACH, B. M., BARTON, S. C., BEECHEY, C. V. & SURANI, M. A. 1991. Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature,* 351**,** 667-70.

FERGUSON-SMITH, A. C., SASAKI, H., CATTANACH, B. M. & SURANI, M. A. 1993. Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature,* 362**,** 751-5.

FIALKOW, P. J. 1970. X-chromosome inactivation and the Xg locus. *Am J Hum Genet,* 22**,** 460-3.

FIRE, A., XU, S., MONTGOMERY, M. K., KOSTAS, S. A., DRIVER, S. E. & MELLO, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature,* 391**,** 806-11.

FOSTER, J. W. & GRAVES, J. A. 1994. An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proc Natl Acad Sci U S A,* 91**,** 1927-31.

FRANKENBERG, S. 2018. Chapter Ten - Pre-gastrula Development of Non-eutherian Mammals. *In:* PLUSA, B. & HADJANTONAKIS, A.-K. (eds.) *Current Topics in Developmental Biology.* Academic Press.

FRANKENBERG, S., SHAW, G., FREYER, C., PASK, A. J. & RENFREE, M. B. 2013. Early cell lineage specification in a marsupial: a case for diverse mechanisms among mammals. *Development,* 140**,** 965-75.

FUJIWARA, Y., KOMIYA, T., KAWABATA, H., SATO, M., FUJIMOTO, H., FURUSAWA, M. & NOCE, T. 1994. Isolation of a DEAD-family protein gene that encodes a murine homolog of Drosophila vasa and its specific expression in germ cell lineage. *Proc Natl Acad Sci U S A,* 91**,** 12258-62.

FULKA, H., MRAZEK, M., TEPLA, O. & FULKA, J., JR. 2004. DNA methylation pattern in human zygotes and developing embryos. *Reproduction,* 128**,** 703-8.

FULKA, J., FULKA, H., SLAVIK, T., OKADA, K. & FULKA, J., JR. 2006. DNA methylation pattern in pig in vivo produced embryos. *Histochem Cell Biol,* 126**,** 213-7.

FUNAKI, S., NAKAMURA, T., NAKATANI, T., UMEHARA, H., NAKASHIMA, H. & NAKANO, T. 2014. Inhibition of maintenance DNA methylation by Stella. *Biochem Biophys Res Commun,* 453**,** 455-60.

GAO, F., NIU, Y., SUN, Y. E., LU, H., CHEN, Y., LI, S., KANG, Y., LUO, Y., SI, C., YU, J., LI, C., SUN, N., SI, W., WANG, H., JI, W. & TAN, T. 2017. De novo DNA methylation during monkey pre-implantation embryogenesis. *Cell Res,* 27**,** 526-539.

GARDINER-GARDEN, M. & FROMMER, M. 1987. CpG islands in vertebrate genomes. *J Mol Biol,* 196**,** 261-82.

GDULA, M. R., NESTEROVA, T. B., PINTACUDA, G., GODWIN, J., ZHAN, Y., OZADAM, H., MCCLELLAN, M., MORALLI, D., KRUEGER, F., GREEN, C. M., REIK, W., KRIAUCIONIS, S., HEARD, E., DEKKER, J. & BROCKDORFF, N. 2019. The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. *Nat Commun,* 10**,** 30.

GENDREL, A. V., APEDAILE, A., COKER, H., TERMANIS, A., ZVETKOVA, I., GODWIN, J., TANG, Y. A., HUNTLEY, D., MONTANA, G., TAYLOR, S., GIANNOULATOU, E., HEARD, E., STANCHEVA, I. & BROCKDORFF, N. 2012. Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev Cell,* 23**,** 265-79.

GIACALONE, J., FRIEDES, J. & FRANCKE, U. 1992. A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat Genet,* 1**,** 137-43.

GINSBURG, M., SNOW, M. H. & MCLAREN, A. 1990. Primordial germ cells in the mouse embryo during gastrulation. *Development,* 110**,** 521-8.

GIORGETTI, L., LAJOIE, B. R., CARTER, A. C., ATTIA, M., ZHAN, Y., XU, J., CHEN, C. J., KAPLAN, N., CHANG, H. Y., HEARD, E. & DEKKER, J. 2016. Structural organization of the inactive X chromosome in the mouse. *Nature,* 535**,** 575-9.

GIRALDEZ, A. J., MISHIMA, Y., RIHEL, J., GROCOCK, R. J., VAN DONGEN, S., INOUE, K., ENRIGHT, A. J. & SCHIER, A. F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science,* 312**,** 75-9.

GKOUNTELA, S., ZHANG, K. X., SHAFIQ, T. A., LIAO, W. W., HARGAN-CALVOPINA, J., CHEN, P. Y. & CLARK, A. T. 2015. DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell,* 161**,** 1425-36.

GLASTAD, K. M., GOKHALE, K., LIEBIG, J. & GOODISMAN, M. A. 2016. The caste- and sex-specific DNA methylome of the termite Zootermopsis nevadensis. *Sci Rep,* 6**,** 37110.

GOLBUS, M. S., CALARCO, P. G. & EPSTEIN, C. J. 1973. The effects of inhibitors of RNA synthesis (alpha-amanitin and actinomycin D) on preimplantation mouse embryogenesis. *J Exp Zool,* 186**,** 207-16.

GOODFELLOW, P. J., MONDELLO, C., DARLING, S. M., PYM, B., LITTLE, P. & GOODFELLOW, P. N. 1988. Absence of methylation of a CpG-rich region at the 5' end of the MIC2 gene on the active X, the inactive X, and the Y chromosome. *Proc Natl Acad Sci U S A,* 85**,** 5605-9.

GORMAN, J. G., DI RE, J., TREACY, A. M. & CAHAN, A. 1963. The application of -Xga antiserum to the question of red cell mosaicism in female heterozygotes. *J Lab Clin Med,* 61**,** 642-9.

GRANT, J., MAHADEVAIAH, S. K., KHIL, P., SANGRITHI, M. N., ROYO, H., DUCKWORTH, J., MCCARREY, J. R., VANDEBERG, J. L., RENFREE, M. B., TAYLOR, W., ELGAR, G., CAMERINI-OTERO, R. D., GILCHRIST, M. J. & TURNER, J. M. 2012. Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature,* 487**,** 254-8.

GRANT, M., ZUCCOTTI, M. & MONK, M. 1992. Methylation of CpG sites of two X-linked genes coincides with X-inactivation in the female mouse embryo but not in the germ line. *Nat Genet,* 2**,** 161-6.

GRAVES, J. A. 1982. 5-azacytidine-induced re-expression of alleles on the inactive X chromosome in a hybrid mouse cell line. *Exp Cell Res,* 141**,** 99-105.

GREALLY, J. M. 2018. A user's guide to the ambiguous word 'epigenetics'. *Nat Rev Mol Cell Biol,* 19**,** 207-208.

GRIPPO, P., IACCARINO, M., PARISI, E. & SCARANO, E. 1968. Methylation of DNA in developing sea urchin embryos. *J Mol Biol,* 36**,** 195-208.

GROSS, D. S. & GARRARD, W. T. 1988. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem,* 57**,** 159-97.

GRUTZNER, F., RENS, W., TSEND-AYUSH, E., EL-MOGHARBEL, N., O'BRIEN, P. C., JONES, R. C., FERGUSON-SMITH, M. A. & MARSHALL GRAVES, J. A. 2004. In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature,* 432**,** 913-7.

GU, H., SMITH, Z. D., BOCK, C., BOYLE, P., GNIRKE, A. & MEISSNER, A. 2011a. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc,* 6**,** 468-81.

GU, T. P., GUO, F., YANG, H., WU, H. P., XU, G. F., LIU, W., XIE, Z. G., SHI, L., HE, X., JIN, S. G., IQBAL, K., SHI, Y. G., DENG, Z., SZABO, P. E., PFEIFER, G. P., LI, J. & XU, G. L. 2011b. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature,* 477**,** 606-10.

GU, Z., EILS, R. & SCHLESNER, M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics,* 32**,** 2847-9.

GUENATRI, M., DUFFIE, R., IRANZO, J., FAUQUE, P. & BOURC'HIS, D. 2013. Plasticity in Dnmt3L-dependent and -independent modes of de novo methylation in the developing mouse embryo. *Development,* 140**,** 562-72.

GUENTHER, M. G., LEVINE, S. S., BOYER, L. A., JAENISCH, R. & YOUNG, R. A. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell,* 130**,** 77-88.

GUIBERT, S., FORNE, T. & WEBER, M. 2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res,* 22**,** 633-41.

GUO, F., LI, X., LIANG, D., LI, T., ZHU, P., GUO, H., WU, X., WEN, L., GU, T. P., HU, B., WALSH, C. P., LI, J., TANG, F. & XU, G. L. 2014a. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell,* 15**,** 447-58.

GUO, F., YAN, L., GUO, H., LI, L., HU, B., ZHAO, Y., YONG, J., HU, Y., WANG, X., WEI, Y., WANG, W., LI, R., YAN, J., ZHI, X., ZHANG, Y., JIN, H., ZHANG, W., HOU, Y., ZHU, P., LI, J., ZHANG, L., LIU, S., REN, Y., ZHU, X., WEN, L., GAO, Y. Q., TANG, F. & QIAO, J. 2015a. The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell,* 161**,** 1437-52.

GUO, H., ZHU, P., GUO, F., LI, X., WU, X., FAN, X., WEN, L. & TANG, F. 2015b. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat Protoc,* 10**,** 645-59.

GUO, H., ZHU, P., WU, X., LI, X., WEN, L. & TANG, F. 2013. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res,* 23**,** 2126-35.

GUO, H., ZHU, P., YAN, L., LI, R., HU, B., LIAN, Y., YAN, J., REN, X., LIN, S., LI, J., JIN, X., SHI, X., LIU, P., WANG, X., WANG, W., WEI, Y., LI, X., GUO, F., WU, X., FAN, X., YONG, J., WEN, L., XIE, S. X., TANG, F. & QIAO, J. 2014b. The DNA methylation landscape of human early embryos. *Nature,* 511**,** 606-610.

GUO, M., ZHANG, Y., ZHOU, J., BI, Y., XU, J., XU, C., KOU, X., ZHAO, Y., LI, Y., TU, Z., LIU, K., LIN, J., YANG, P., GAO, S. & WANG, Y. 2019. Precise temporal regulation of Dux is important for embryo development. *Cell Res,* 29**,** 956-959.

GUPTA, V., PARISI, M., STURGILL, D., NUTTALL, R., DOCTOLERO, M., DUDKO, O. K., MALLEY, J. D., EASTMAN, P. S. & OLIVER, B. 2006. Global analysis of X-chromosome dosage compensation. *J Biol,* 5**,** 3.

HACKETT, J. A., REDDINGTON, J. P., NESTOR, C. E., DUNICAN, D. S., BRANCO, M. R., REICHMANN, J., REIK, W., SURANI, M. A., ADAMS, I. R. & MEEHAN, R. R. 2012. Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development,* 139**,** 3623-32.

HACKETT, JAMIE A. & SURANI, M. A. 2013. Beyond DNA: Programming and Inheritance of Parental Methylomes. *Cell,* 153**,** 737-739.

HAHNE, F. & IVANEK, R. 2016. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol,* 1418**,** 335-51.

HAJKOVA, P., ANCELIN, K., WALDMANN, T., LACOSTE, N., LANGE, U. C., CESARI, F., LEE, C., ALMOUZNI, G., SCHNEIDER, R. & SURANI, M. A. 2008. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature,* 452**,** 877-81.

HAJKOVA, P., ERHARDT, S., LANE, N., HAAF, T., EL-MAARRI, O., REIK, W., WALTER, J. & SURANI, M. A. 2002. Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev,* 117**,** 15-23.

HAJKOVA, P., JEFFRIES, S. J., LEE, C., MILLER, N., JACKSON, S. P. & SURANI, M. A. 2010. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science,* 329**,** 78-82.

HAMATANI, T., CARTER, M. G., SHAROV, A. A. & KO, M. S. 2004. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell,* 6**,** 117-31.

HAMMOND, S. M., BERNSTEIN, E., BEACH, D. & HANNON, G. J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature,* 404**,** 293-6.

HAN, L., REN, C., ZHANG, J., SHU, W. & WANG, Q. 2019. Differential roles of Stella in the modulation of DNA methylation during oocyte and zygotic development. *Cell Discov,* 5**,** 9.

HANNA, C. W., PEREZ-PALACIOS, R., GAHUROVA, L., SCHUBERT, M., KRUEGER, F., BIGGINS, L., ANDREWS, S., COLOME-TATCHE, M., BOURC'HIS, D., DEAN, W. & KELSEY, G. 2019. Endogenous retroviral insertions drive non-canonical imprinting in extra-embryonic tissues. *Genome Biol,* 20**,** 225.

HARRY, J. L., KOOPMAN, P., BRENNAN, F. E., GRAVES, J. A. & RENFREE, M. B. 1995. Widespread expression of the testis-determining gene SRY in a marsupial. *Nat Genet,* 11**,** 347-9.

HATA, K., OKANO, M., LEI, H. & LI, E. 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development,* 129**,** 1983-93.

HE, Y. F., LI, B. Z., LI, Z., LIU, P., WANG, Y., TANG, Q., DING, J., JIA, Y., CHEN, Z., LI, L., SUN, Y., LI, X., DAI, Q., SONG, C. X., ZHANG, K., HE, C. & XU, G. L. 2011.

Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science,* 333**,** 1303-7.
HEARD, E., ROUGEULLE, C., ARNAUD, D., AVNER, P., ALLIS, C. D. & SPECTOR, D. L. 2001. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell,* 107**,** 727-38.
HELLMAN, A. & CHESS, A. 2007. Gene body-specific methylation on the active X chromosome. *Science,* 315**,** 1141-3.
HENDRICKSON, P. G., DORAIS, J. A., GROW, E. J., WHIDDON, J. L., LIM, J. W., WIKE, C. L., WEAVER, B. D., PFLUEGER, C., EMERY, B. R., WILCOX, A. L., NIX, D. A., PETERSON, C. M., TAPSCOTT, S. J., CARRELL, D. T. & CAIRNS, B. R. 2017. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet,* 49**,** 925-934.
HICKFORD, D. E., FRANKENBERG, S., PASK, A. J., SHAW, G. & RENFREE, M. B. 2011. DDX4 (VASA) is conserved in germ cell development in marsupials and monotremes. *Biol Reprod,* 85**,** 733-43.
HILL, P. W. S., LEITCH, H. G., REQUENA, C. E., SUN, Z., AMOUROUX, R., ROMAN-TRUFERO, M., BORKOWSKA, M., TERRAGNI, J., VAISVILA, R., LINNETT, S., BAGCI, H., DHARMALINGHAM, G., HABERLE, V., LENHARD, B., ZHENG, Y., PRADHAN, S. & HAJKOVA, P. 2018. Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature,* 555**,** 392-396.
HIRASAWA, R., CHIBA, H., KANEDA, M., TAJIMA, S., LI, E., JAENISCH, R. & SASAKI, H. 2008. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev,* 22**,** 1607-16.
HODGES, E., MOLARO, A., DOS SANTOS, C. O., THEKKAT, P., SONG, Q., UREN, P. J., PARK, J., BUTLER, J., RAFII, S., MCCOMBIE, W. R., SMITH, A. D. & HANNON, G. J. 2011. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell,* 44**,** 17-28.
HOLLIDAY, R. & GRIGG, G. W. 1993. DNA methylation and mutation. *Mutat Res,* 285**,** 61-7.
HOLLIDAY, R. & PUGH, J. E. 1975. DNA modification mechanisms and gene activity during development. *Science,* 187**,** 226-32.
HOLOCH, D. & MOAZED, D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet,* 16**,** 71-84.
HON, G. C., RAJAGOPAL, N., SHEN, Y., MCCLEARY, D. F., YUE, F., DANG, M. D. & REN, B. 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet,* 45**,** 1198-206.
HORE, T. A., KOINA, E., WAKEFIELD, M. J. & MARSHALL GRAVES, J. A. 2007. The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals. *Chromosome Res,* 15**,** 147-61.
HOU, J., LEI, T. H., LIU, L., CUI, X. H., AN, X. R. & CHEN, Y. F. 2005. DNA methylation patterns in in vitro-fertilised goat zygotes. *Reprod Fertil Dev,* 17**,** 809-13.
HOU, J., LIU, L., ZHANG, J., CUI, X. H., YAN, F. X., GUAN, H., CHEN, Y. F. & AN, X. R. 2008. Epigenetic modification of histone 3 at lysine 9 in sheep zygotes and its relationship with DNA methylation. *BMC Dev Biol,* 8**,** 60.
HOWELL, C. Y., BESTOR, T. H., DING, F., LATHAM, K. E., MERTINEIT, C., TRASLER, J. M. & CHAILLET, J. R. 2001. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell,* 104**,** 829-38.
HOWLETT, S. K. & REIK, W. 1991. Methylation levels of maternal and paternal genomes during preimplantation development. *Development,* 113**,** 119-127.
HUGHES, J. F., SKALETSKY, H., KOUTSEVA, N., PYNTIKOVA, T. & PAGE, D. C. 2015. Sex chromosome-to-autosome transposition events counter Y-chromosome gene loss in mammals. *Genome Biol,* 16**,** 104.

HUNTRISS, J., HINKINS, M., OLIVER, B., HARRIS, S. E., BEAZLEY, J. C., RUTHERFORD, A. J., GOSDEN, R. G., LANZENDORF, S. E. & PICTON, H. M. 2004. Expression of mRNAs for DNA methyltransferases and methyl-CpG-binding proteins in the human female germ line, preimplantation embryos, and embryonic stem cells. *Mol Reprod Dev,* 67**,** 323-36.

ILLINGWORTH, R. S. & BIRD, A. P. 2009. CpG islands--'a rough guide'. *FEBS Lett,* 583**,** 1713-20.

INOUE, A., JIANG, L., LU, F., SUZUKI, T. & ZHANG, Y. 2017a. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature,* 547**,** 419-424.

INOUE, A., JIANG, L., LU, F. & ZHANG, Y. 2017b. Genomic imprinting of Xist by maternal H3K27me3. *Genes Dev,* 31**,** 1927-1932.

INOUE, A., SHEN, L., DAI, Q., HE, C. & ZHANG, Y. 2011. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res,* 21**,** 1670-6.

INOUE, A., SHEN, L., MATOBA, S. & ZHANG, Y. 2015. Haploinsufficiency, but not defective paternal 5mC oxidation, accounts for the developmental defects of maternal Tet3 knockouts. *Cell Rep,* 10**,** 463-70.

IQBAL, K., JIN, S. G., PFEIFER, G. P. & SZABO, P. E. 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A,* 108**,** 3642-7.

IRIZARRY, R. A., LADD-ACOSTA, C., WEN, B., WU, Z., MONTANO, C., ONYANGO, P., CUI, H., GABO, K., RONGIONE, M. & WEBSTER, M. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet,* 41.

ISHIHARA, T., HICKFORD, D., SHAW, G., PASK, A. J. & RENFREE, M. B. 2019. DNA methylation dynamics in the germline of the marsupial tammar wallaby, Macropus eugenii. *DNA Res,* 26**,** 85-94.

ITO, S., D'ALESSIO, A. C., TARANOVA, O. V., HONG, K., SOWERS, L. C. & ZHANG, Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature,* 466**,** 1129-33.

ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-3.

JEGALIAN, K. & PAGE, D. C. 1998. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature,* 394**,** 776-80.

JENUWEIN, T. & ALLIS, C. D. 2001. Translating the histone code. *Science,* 293**,** 1074-80.

JEONG, M., SUN, D., LUO, M., HUANG, Y., CHALLEN, G. A., RODRIGUEZ, B., ZHANG, X., CHAVEZ, L., WANG, H., HANNAH, R., KIM, S. B., YANG, L., KO, M., CHEN, R., GOTTGENS, B., LEE, J. S., GUNARATNE, P., GODLEY, L. A., DARLINGTON, G. J., RAO, A., LI, W. & GOODELL, M. A. 2014. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet,* 46**,** 17-23.

JEONG, Y. S., YEO, S., PARK, J. S., KOO, D. B., CHANG, W. K., LEE, K. K. & KANG, Y. K. 2007. DNA methylation state is preserved in the sperm-derived pronucleus of the pig zygote. *Int J Dev Biol,* 51**,** 707-14.

JEPPESEN, P. & TURNER, B. M. 1993. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell,* 74**,** 281-9.

JIANG, L., ZHANG, J., WANG, J. J., WANG, L., ZHANG, L., LI, G., YANG, X., MA, X., SUN, X., CAI, J., ZHANG, J., HUANG, X., YU, M., WANG, X., LIU, F., WU, C. I., HE, C., ZHANG, B., CI, W. & LIU, J. 2013. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell,* 153**,** 773-84.

JIANG, Z., LIN, J., DONG, H., ZHENG, X., MARJANI, S. L., DUAN, J., OUYANG, Z., CHEN, J. & TIAN, X. C. 2018. DNA methylomes of bovine gametes and in vivo produced preimplantation embryos. *Biol Reprod,* 99**,** 949-959.

JONES, P. A. & TAYLOR, S. M. 1980. Cellular differentiation, cytidine analogs and DNA methylation. *Cell,* 20**,** 85-93.

JOO, J. E., NOVAKOVIC, B., CRUICKSHANK, M., DOYLE, L. W., CRAIG, J. M. & SAFFERY, R. 2014. Human active X-specific DNA methylation events showing stability across time and tissues. *Eur J Hum Genet,* 22**,** 1376-81.

JULIEN, P., BRAWAND, D., SOUMILLON, M., NECSULEA, A., LIECHTI, A., SCHUTZ, F., DAISH, T., GRUTZNER, F. & KAESSMANN, H. 2012. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol,* 10**,** e1001328.

KAFRI, T., ARIEL, M., BRANDEIS, M., SHEMER, R., URVEN, L., MCCARREY, J., CEDAR, H. & RAZIN, A. 1992. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes & Development,* 6**,** 705-714.

KAGIWADA, S., KURIMOTO, K., HIROTA, T., YAMAJI, M. & SAITOU, M. 2013. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J,* 32**,** 340-53.

KANE, D. A., HAMMERSCHMIDT, M., MULLINS, M. C., MAISCHEIN, H. M., BRAND, M., VAN EEDEN, F. J., FURUTANI-SEIKI, M., GRANATO, M., HAFFTER, P., HEISENBERG, C. P., JIANG, Y. J., KELSH, R. N., ODENTHAL, J., WARGA, R. M. & NUSSLEIN-VOLHARD, C. 1996. The zebrafish epiboly mutants. *Development,* 123**,** 47-55.

KANE, D. A. & KIMMEL, C. B. 1993. The zebrafish midblastula transition. *Development,* 119**,** 447-56.

KANEDA, M., OKANO, M., HATA, K., SADO, T., TSUJIMOTO, N., LI, E. & SASAKI, H. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature,* 429**,** 900-3.

KANG, J., LIENHARD, M., PASTOR, W. A., CHAWLA, A., NOVOTNY, M., TSAGARATOU, A., LASKEN, R. S., THOMPSON, E. C., SURANI, M. A., KORALOV, S. B., KALANTRY, S., CHAVEZ, L. & RAO, A. 2015. Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc Natl Acad Sci U S A,* 112**,** E4236-45.

KASLOW, D. C. & MIGEON, B. R. 1987. DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. *Proc Natl Acad Sci U S A,* 84**,** 6210-4.

KAZACHENKA, A., BERTOZZI, T. M., SJOBERG-HERRERA, M. K., WALKER, N., GARDNER, J., GUNNING, R., PAHITA, E., ADAMS, S., ADAMS, D. & FERGUSON-SMITH, A. C. 2018. Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell,* 175**,** 1717.

KEANE, T. M., GOODSTADT, L., DANECEK, P., WHITE, M. A., WONG, K., YALCIN, B., HEGER, A., AGAM, A., SLATER, G., GOODSON, M., FURLOTTE, N. A., ESKIN, E., NELLAKER, C., WHITLEY, H., CLEAK, J., JANOWITZ, D., HERNANDEZ-PLIEGO, P., EDWARDS, A., BELGARD, T. G., OLIVER, P. L., MCINTYRE, R. E., BHOMRA, A., NICOD, J., GAN, X., YUAN, W., VAN DER WEYDEN, L., STEWARD, C. A., BALA, S., STALKER, J., MOTT, R., DURBIN, R., JACKSON, I. J., CZECHANSKI, A., GUERRA-ASSUNCAO, J. A., DONAHUE, L. R., REINHOLDT, L. G., PAYSEUR, B. A., PONTING, C. P., BIRNEY, E., FLINT, J. & ADAMS, D. J. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature,* 477**,** 289-94.

KEITH, D. H., SINGER-SAM, J. & RIGGS, A. D. 1986. Active X chromosome DNA is unmethylated at eight CCGG sites clustered in a guanine-plus-cytosine-rich island at the 5' end of the gene for phosphoglycerate kinase. *Mol Cell Biol,* 6**,** 4122-5.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KEOWN, C. L., BERLETCH, J. B., CASTANON, R., NERY, J. R., DISTECHE, C. M., ECKER, J. R. & MUKAMEL, E. A. 2017. Allele-specific non-CG DNA methylation marks domains of active chromatin in female mouse brain. *Proc Natl Acad Sci U S A,* 114**,** E2882-E2890.

KERMICLE, J. L. 1970. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics,* 66**,** 69-85.

KHALIL, A. M., GUTTMAN, M., HUARTE, M., GARBER, M., RAJ, A., RIVEA MORALES, D., THOMAS, K., PRESSER, A., BERNSTEIN, B. E., VAN OUDENAARDEN, A., REGEV, A., LANDER, E. S. & RINN, J. L. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A,* 106**,** 11667-72.

KIGAMI, D., MINAMI, N., TAKAYAMA, H. & IMAI, H. 2003. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol Reprod,* 68**,** 651-4.

KILLIAN, J. K., BYRD, J. C., JIRTLE, J. V., MUNDAY, B. L., STOSKOPF, M. K., MACDONALD, R. G. & JIRTLE, R. L. 2000. M6P/IGF2R imprinting evolution in mammals. *Mol Cell,* 5**,** 707-16.

KIM, D., PAGGI, J. M., PARK, C., BENNETT, C. & SALZBERG, S. L. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol,* 37**,** 907-915.

KIM, T. K., HEMBERG, M., GRAY, J. M., COSTA, A. M., BEAR, D. M., WU, J., HARMIN, D. A., LAPTEWICZ, M., BARBARA-HALEY, K., KUERSTEN, S., MARKENSCOFF-PAPADIMITRIOU, E., KUHL, D., BITO, H., WORLEY, P. F., KREIMAN, G. & GREENBERG, M. E. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature,* 465**,** 182-7.

KIMMEL, C. B., BALLARD, W. W., KIMMEL, S. R., ULLMANN, B. & SCHILLING, T. F. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn,* 203**,** 253-310.

KISHIGAMI, S., VAN THUAN, N., HIKICHI, T., OHTA, H., WAKAYAMA, S., MIZUTANI, E. & WAKAYAMA, T. 2006. Epigenetic abnormalities of the mouse paternal zygotic genome associated with microinsemination of round spermatids. *Dev Biol,* 289**,** 195-205.

KO, M., BANDUKWALA, H. S., AN, J., LAMPERTI, E. D., THOMPSON, E. C., HASTIE, R., TSANGARATOU, A., RAJEWSKY, K., KORALOV, S. B. & RAO, A. 2011. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc Natl Acad Sci U S A,* 108**,** 14566-71.

KOBAYASHI, H., SAKURAI, T., IMAI, M., TAKAHASHI, N., FUKUDA, A., YAYOI, O., SATO, S., NAKABAYASHI, K., HATA, K., SOTOMARU, Y., SUZUKI, Y. & KONO, T. 2012. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet,* 8**,** e1002440.

KOBOLDT, D. C., CHEN, K., WYLIE, T., LARSON, D. E., MCLELLAN, M. D., MARDIS, E. R., WEINSTOCK, G. M., WILSON, R. K. & DING, L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics,* 25**,** 2283-5.

KOINA, E., CHAUMEIL, J., GREAVES, I. K., TREMETHICK, D. J. & GRAVES, J. A. 2009. Specific patterns of histone marks accompany X chromosome inactivation in a marsupial. *Chromosome Res,* 17**,** 115-26.

KONO, T., OBATA, Y., YOSHIMZU, T., NAKAHARA, T. & CARROLL, J. 1996. Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. *Nat Genet,* 13**,** 91-4.

KOOPMAN, P., GUBBAY, J., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1991. Male development of chromosomally female mice transgenic for Sry. *Nature,* 351**,** 117-21.

KORNBERG, R. D. & THOMAS, J. O. 1974. Chromatin structure; oligomers of the histones. *Science,* 184**,** 865-8.

KRIBELBAUER, J. F., LAPTENKO, O., CHEN, S., MARTINI, G. D., FREED-PASTOR, W. A., PRIVES, C., MANN, R. S. & BUSSEMAKER, H. J. 2017. Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. *Cell Rep,* 19**,** 2383-2395.

KRUEGER, F. 2012. *TrimGalore! A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries.* [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [Accessed].

KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics,* 27**,** 1571-2.

KRUEGER, F. & ANDREWS, S. R. 2016. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res,* 5**,** 1479.

KUO, H. C., LIN, P. Y., CHUNG, T. C., CHAO, C. M., LAI, L. C., TSAI, M. H. & CHUANG, E. Y. 2011. DBCAT: database of CpG islands and analytical tools for identifying comprehensive methylation profiles in cancer cells. *J Comput Biol,* 18**,** 1013-7.

KURIHARA, Y., KAWAMURA, Y., UCHIJIMA, Y., AMAMO, T., KOBAYASHI, H., ASANO, T. & KURIHARA, H. 2008. Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. *Dev Biol,* 313**,** 335-46.

LA SALLE, S. & TRASLER, J. M. 2006. Dynamic expression of DNMT3a and DNMT3b isoforms during male germ cell development in the mouse. *Dev Biol,* 296**,** 71-82.

LAHN, B. T. & PAGE, D. C. 1999. Four evolutionary strata on the human X chromosome. *Science,* 286**,** 964-7.

LAI, F., OROM, U. A., CESARONI, M., BERINGER, M., TAATJES, D. J., BLOBEL, G. A. & SHIEKHATTAR, R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature,* 494**,** 497-501.

LANE, N., DEAN, W., ERHARDT, S., HAJKOVA, P., SURANI, A., WALTER, J. & REIK, W. 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis,* 35**,** 88-93.

LARSEN, F., GUNDERSEN, G., LOPEZ, R. & PRYDZ, H. 1992. CpG islands as gene markers in the human genome. *Genomics,* 13**,** 1095-107.

LATOS, P. A., PAULER, F. M., KOERNER, M. V., SENERGIN, H. B., HUDSON, Q. J., STOCSITS, R. R., ALLHOFF, W., STRICKER, S. H., KLEMENT, R. M., WARCZOK, K. E., AUMAYR, K., PASIERBEK, P. & BARLOW, D. P. 2012. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science,* 338**,** 1469-72.

LAWRENCE, M., GENTLEMAN, R. & CAREY, V. 2009. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics,* 25**,** 1841-2.

LAWRENCE, M., HUBER, W., PAGES, H., ABOYOUN, P., CARLSON, M., GENTLEMAN, R., MORGAN, M. T. & CAREY, V. J. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol,* 9**,** e1003118.

LAWSON, K. A. & HAGE, W. J. 1994. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp,* 182**,** 68-84; discussion 84-91.

LEE, J. T. 2000. Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. *Cell,* 103**,** 17-27.

LEE, J. T., DAVIDOW, L. S. & WARSHAWSKY, D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet,* 21**,** 400-4.
LEE, M. T., BONNEAU, A. R. & GIRALDEZ, A. J. 2014a. Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol,* 30**,** 581-613.
LEE, M. T., BONNEAU, A. R., TAKACS, C. M., BAZZINI, A. A., DIVITO, K. R., FLEMING, E. S. & GIRALDEZ, A. J. 2013. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature,* 503**,** 360-4.
LEE, Y. K., JIN, S., DUAN, S., LIM, Y. C., NG, D. P., LIN, X. M., YEO, G. S. H. & DING, C. 2014b. Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. *Biological Procedures Online,* 16.
LEPIKHOV, K., ZAKHARTCHENKO, V., HAO, R., YANG, F., WRENZYCKI, C., NIEMANN, H., WOLF, E. & WALTER, J. 2008. Evidence for conserved DNA and histone H3 methylation reprogramming in mouse, bovine and rabbit zygotes. *Epigenetics Chromatin,* 1**,** 8.
LEWIS, J. D., MEEHAN, R. R., HENZEL, W. J., MAURER-FOGY, I., JEPPESEN, P., KLEIN, F. & BIRD, A. 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell,* 69**,** 905-14.
LI, E., BEARD, C. & JAENISCH, R. 1993. Role for DNA methylation in genomic imprinting. *Nature,* 366**,** 362-5.
LI, E., BESTOR, T. H. & JAENISCH, R. 1992a. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69**,** 915-26.
LI, E., BESTOR, T. H. & JAENISCH, R. 1992b. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69.
LI, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv e-prints* [Online]. Available: https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L [Accessed March 01, 2013].
LI H., HANDSAKER B., WYSOKER A., FENNELL T., RUAN J., HOMER N., MARTH G., ABECASIS G., DURBIN R. & 1000 GENOME PROJECT DATA PROCESSING SUBGROUP . 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.
LI, X., ITO, M., ZHOU, F., YOUNGSON, N., ZUO, X., LEDER, P. & FERGUSON-SMITH, A. C. 2008. A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev Cell,* 15**,** 547-57.
LI, Y., ZHANG, Z., CHEN, J., LIU, W., LAI, W., LIU, B., LI, X., LIU, L., XU, S., DONG, Q., WANG, M., DUAN, X., TAN, J., ZHENG, Y., ZHANG, P., FAN, G., WONG, J., XU, G. L., WANG, Z., WANG, H., GAO, S. & ZHU, B. 2018. Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature,* 564**,** 136-140.
LI, Z., CAI, X., CAI, C. L., WANG, J., ZHANG, W., PETERSEN, B. E., YANG, F. C. & XU, M. 2011. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood,* 118**,** 4509-18.
LIEBERMAN-AIDEN, E., VAN BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J., DORSCHNER, M. O., SANDSTROM, R., BERNSTEIN, B., BENDER, M. A., GROUDINE, M., GNIRKE, A., STAMATOYANNOPOULOS, J., MIRNY, L. A., LANDER, E. S. & DEKKER, J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science,* 326**,** 289-93.
LIN, H., HALSALL, J. A., ANTCZAK, P., O'NEILL, L. P., FALCIANI, F. & TURNER, B. M. 2011. Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis. *Nat Genet,* 43**,** 1169-70; author reply 1171-2.
LINDSAY, S., MONK, M., HOLLIDAY, R., HUSCHTSCHA, L., DAVIES, K. E., RIGGS, A. D. & FLAVELL, R. A. 1985. Differences in methylation on the active and inactive human X chromosomes. *Ann Hum Genet,* 49**,** 115-27.

LISTER, R., MUKAMEL, E. A., NERY, J. R., URICH, M., PUDDIFOOT, C. A., JOHNSON, N. D., LUCERO, J., HUANG, Y., DWORK, A. J., SCHULTZ, M. D., YU, M., TONTI-FILIPPINI, J., HEYN, H., HU, S., WU, J. C., RAO, A., ESTELLER, M., HE, C., HAGHIGHI, F. G., SEJNOWSKI, T. J., BEHRENS, M. M. & ECKER, J. R. 2013. Global epigenomic reconfiguration during mammalian brain development. *Science,* 341**,** 1237905.

LISTER, R., O'MALLEY, R. C., TONTI-FILIPPINI, J., GREGORY, B. D., BERRY, C. C., MILLAR, A. H. & ECKER, J. R. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell,* 133**,** 523-36.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z. & NGO, Q. M. 2009a. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009b. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.

LOCK, L. F., MELTON, D. W., CASKEY, C. T. & MARTIN, G. R. 1986. Methylation of the mouse hprt gene differs on the active and inactive X chromosomes. *Mol Cell Biol,* 6**,** 914-24.

LOEBEL, D. A. & JOHNSTON, P. G. 1993. Analysis of DNase 1 sensitivity and methylation of active and inactive X chromosomes of kangaroos (Macropus robustus) by in situ nick translation. *Chromosoma,* 102**,** 81-7.

LOEBEL, D. A. & JOHNSTON, P. G. 1996. Methylation analysis of a marsupial X-linked CpG island by bisulfite genomic sequencing. *Genome Res,* 6**,** 114-23.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.

LUO, Z. X., YUAN, C. X., MENG, Q. J. & JI, Q. 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature,* 476**,** 442-5.

LYON, M. F. 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature,* 190**,** 372-3.

MAATOUK, D. M., KELLAM, L. D., MANN, M. R., LEI, H., LI, E., BARTOLOMEI, M. S. & RESNICK, J. L. 2006. DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development,* 133**,** 3411-8.

MACFARLAN, T. S., GIFFORD, W. D., AGARWAL, S., DRISCOLL, S., LETTIERI, K., WANG, J., ANDREWS, S. E., FRANCO, L., ROSENFELD, M. G., REN, B. & PFAFF, S. L. 2011. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev,* 25**,** 594-607.

MACFARLAN, T. S., GIFFORD, W. D., DRISCOLL, S., LETTIERI, K., ROWE, H. M., BONANOMI, D., FIRTH, A., SINGER, O., TRONO, D. & PFAFF, S. L. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature,* 487**,** 57-63.

MACLARY, E., BUTTIGIEG, E., HINTEN, M., GAYEN, S., HARRIS, C., SARKAR, M. K., PURUSHOTHAMAN, S. & KALANTRY, S. 2014. Differentiation-dependent requirement of Tsix long non-coding RNA in imprinted X-chromosome inactivation. *Nat Commun,* 5**,** 4209.

MACLEOD, D., CLARK, V. H. & BIRD, A. 1999. Absence of genome-wide changes in DNA methylation during development of the zebrafish. *Nat Genet,* 23**,** 139-140.

MAENOHARA, S., UNOKI, M., TOH, H., OHISHI, H., SHARIF, J., KOSEKI, H. & SASAKI, H. 2017. Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. *PLoS Genet,* 13**,** e1007042.

MAHADEVAIAH, S. K., ROYO, H., VANDEBERG, J. L., MCCARREY, J. R., MACKAY, S. & TURNER, J. M. 2009. Key features of the X inactivation process are conserved between marsupials and eutherians. *Curr Biol,* 19**,** 1478-84.

MAK, W., NESTEROVA, T. B., DE NAPOLES, M., APPANAH, R., YAMANAKA, S., OTTE, A. P. & BROCKDORFF, N. 2004. Reactivation of the paternal X chromosome in early mouse embryos. *Science,* 303**,** 666-9.

MARAHRENS, Y., PANNING, B., DAUSMAN, J., STRAUSS, W. & JAENISCH, R. 1997. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev,* 11**,** 156-66.

MARTINEZ-ARGUELLES, D. B., LEE, S. & PAPADOPOULOS, V. 2014. In silico analysis identifies novel restriction enzyme combinations that expand reduced representation bisulfite sequencing CpG coverage. *BMC research notes,* 7.

MARUSHIGE, Y. & MARUSHIGE, K. 1975. Transformation of sperm histone during formation and maturation of rat spermatozoa. *J Biol Chem,* 250**,** 39-45.

MASALA, L., BURRAI, G. P., BELLU, E., ARIU, F., BOGLIOLO, L., LEDDA, S. & BEBBERE, D. 2017. Methylation dynamics during folliculogenesis and early embryo development in sheep. *Reproduction,* 153**,** 605-619.

MATE, K. E., ROBINSON, E. S., VANDEBERG, J. L. & PEDERSEN, R. A. 1994. Timetable of in vivo embryonic development in the grey short-tailed opossum (Monodelphis domestica). *Mol Reprod Dev,* 39**,** 365-74.

MAUNAKEA, A. K., CHEPELEV, I., CUI, K. & ZHAO, K. 2013. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res,* 23**,** 1256-69.

MAYER, W., NIVELEAU, A., WALTER, J., FUNDELE, R. & HAAF, T. 2000. Demethylation of the zygotic paternal genome. *Nature,* 403**,** 501-502.

MCGHEE, J. D. & GINDER, G. D. 1979. Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature,* 280**,** 419-420.

MCGRATH, J. & SOLTER, D. 1983. Nuclear transplantation in mouse embryos. *J Exp Zool,* 228**,** 355-62.

MCGRAW, S., OAKES, C. C., MARTEL, J., CIRIO, M. C., DE ZEEUW, P., MAK, W., PLASS, C., BARTOLOMEI, M. S., CHAILLET, J. R. & TRASLER, J. M. 2013. Loss of DNMT1o disrupts imprinted X chromosome inactivation and accentuates placental defects in females. *PLoS Genet,* 9**,** e1003873.

MEEHAN, R. R., LEWIS, J. D., MCKAY, S., KLEINER, E. L. & BIRD, A. P. 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell,* 58**,** 499-507.

MEISSNER, A., MIKKELSEN, T. S., GU, H., WERNIG, M., HANNA, J., SIVACHENKO, A., ZHANG, X., BERNSTEIN, B. E., NUSBAUM, C., JAFFE, D. B., GNIRKE, A., JAENISCH, R. & LANDER, E. S. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature,* 454**,** 766-70.

MIGEON, B. R., JAN DE BEUR, S. & AXELMAN, J. 1989. Frequent derepression of G6PD and HPRT on the marsupial inactive X chromosome associated with cell proliferation in vitro. *Exp Cell Res,* 182**,** 597-609.

MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCHE, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007a. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature,* 448**,** 553-60.

MIKKELSEN, T. S., WAKEFIELD, M. J., AKEN, B., AMEMIYA, C. T., CHANG, J. L., DUKE, S., GARBER, M., GENTLES, A. J., GOODSTADT, L., HEGER, A., JURKA, J., KAMAL, M., MAUCELI, E., SEARLE, S. M., SHARPE, T., BAKER, M. L., BATZER, M. A., BENOS, P. V., BELOV, K., CLAMP, M., COOK, A., CUFF, J.,

DAS, R., DAVIDOW, L., DEAKIN, J. E., FAZZARI, M. J., GLASS, J. L., GRABHERR, M., GREALLY, J. M., GU, W., HORE, T. A., HUTTLEY, G. A., KLEBER, M., JIRTLE, R. L., KOINA, E., LEE, J. T., MAHONY, S., MARRA, M. A., MILLER, R. D., NICHOLLS, R. D., ODA, M., PAPENFUSS, A. T., PARRA, Z. E., POLLOCK, D. D., RAY, D. A., SCHEIN, J. E., SPEED, T. P., THOMPSON, K., VANDEBERG, J. L., WADE, C. M., WALKER, J. A., WATERS, P. D., WEBBER, C., WEIDMAN, J. R., XIE, X., ZODY, M. C., BROAD INSTITUTE GENOME SEQUENCING, P., BROAD INSTITUTE WHOLE GENOME ASSEMBLY, T., GRAVES, J. A., PONTING, C. P., BREEN, M., SAMOLLOW, P. B., LANDER, E. S. & LINDBLAD-TOH, K. 2007b. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature,* 447**,** 167-77.

MILLER, D. A., OKAMOTO, E., ERLANGER, B. F. & MILLER, O. J. 1982. Is DNA methylation responsible for mammalian X chromosome inactivation? *Cytogenet Cell Genet,* 33**,** 345-9.

MOHANDAS, T., SPARKES, R. S. & SHAPIRO, L. J. 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science,* 211**,** 393-6.

MOLARO, A., FALCIATORI, I., HODGES, E., ARAVIN, A. A., MARRAN, K., RAFII, S., MCCOMBIE, W. R., SMITH, A. D. & HANNON, G. J. 2014. Two waves of de novo methylation during mouse germ cell development. *Genes Dev,* 28**,** 1544-9.

MONK, M., BOUBELIK, M. & LEHNERT, S. 1987. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development,* 99**,** 371-382.

MONK, M. & HARPER, M. 1978. X-chromosome activity in preimplantation mouse embryos from XX and XO mothers. *J Embryol Exp Morphol,* 46**,** 53-64.

MOORE, T. & HAIG, D. 1991. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet,* 7**,** 45-9.

MORAN-CRUSIO, K., REAVIE, L., SHIH, A., ABDEL-WAHAB, O., NDIAYE-LOBRY, D., LOBRY, C., FIGUEROA, M. E., VASANTHAKUMAR, A., PATEL, J., ZHAO, X., PERNA, F., PANDEY, S., MADZO, J., SONG, C., DAI, Q., HE, C., IBRAHIM, S., BERAN, M., ZAVADIL, J., NIMER, S. D., MELNICK, A., GODLEY, L. A., AIFANTIS, I. & LEVINE, R. L. 2011. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell,* 20**,** 11-24.

MORGAN, H. D., SUTHERLAND, H. G., MARTIN, D. I. & WHITELAW, E. 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet,* 23**,** 314-8.

MORSELLI, M., PASTOR, W. A., MONTANINI, B., NEE, K., FERRARI, R., FU, K., BONORA, G., RUBBI, L., CLARK, A. T., OTTONELLO, S., JACOBSEN, S. E. & PELLEGRINI, M. 2015. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *Elife,* 4**,** e06205.

MULHOLLAND, C. B., NISHIYAMA, A., RYAN, J., NAKAMURA, R., YIĞIT, M., GLÜCK, I. M., TRUMMER, C., QIN, W., BARTOSCHEK, M. D., TRAUBE, F. R., PARSA, E., UGUR, E., MODIC, M., ACHARYA, A., STOLZ, P., ZIEGENHAIN, C., WIERER, M., ENARD, W., CARELL, T., LAMB, D. C., TAKEDA, H., NAKANASHI, M., BULTMANN, S. & LEONHARDT, H. 2020. Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive demethylation in mammals. *bioRxiv***,** 321604.

NAKAMURA, T., ARAI, Y., UMEHARA, H., MASUHARA, M., KIMURA, T., TANIGUCHI, H., SEKIMOTO, T., IKAWA, M., YONEDA, Y., OKABE, M., TANAKA, S., SHIOTA, K. & NAKANO, T. 2007. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat Cell Biol,* 9**,** 64-71.

NAKAMURA, T., LIU, Y. J., NAKASHIMA, H., UMEHARA, H., INOUE, K., MATOBA, S., TACHIBANA, M., OGURA, A., SHINKAI, Y. & NAKANO, T. 2012. PGC7 binds

histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature,* 486**,** 415-9.
NAN, X., NG, H. H., JOHNSON, C. A., LAHERTY, C. D., TURNER, B. M., EISENMAN, R. N. & BIRD, A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature,* 393**,** 386-9.
NARASIMHAN, V., DANECEK, P., SCALLY, A., XUE, Y., TYLER-SMITH, C. & DURBIN, R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics,* 32**,** 1749-51.
NAVEH-MANY, T. & CEDAR, H. 1981. Active gene sequences are undermethylated. *Proc Natl Acad Sci U S A,* 78**,** 4246-50.
NERI, F., RAPELLI, S., KREPELOVA, A., INCARNATO, D., PARLATO, C., BASILE, G., MALDOTTI, M., ANSELMI, F. & OLIVIERO, S. 2017. Intragenic DNA methylation prevents spurious transcription initiation. *Nature,* 543**,** 72-77.
NG, H. H., ZHANG, Y., HENDRICH, B., JOHNSON, C. A., TURNER, B. M., ERDJUMENT-BROMAGE, H., TEMPST, P., REINBERG, D. & BIRD, A. 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet,* 23**,** 58-61.
NG, R. K., DEAN, W., DAWSON, C., LUCIFERO, D., MADEJA, Z., REIK, W. & HEMBERGER, M. 2008. Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol,* 10**,** 1280-90.
NGUYEN, D. K. & DISTECHE, C. M. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet,* 38**,** 47-53.
NIAKAN, K. K., HAN, J., PEDERSEN, R. A., SIMON, C. & PERA, R. A. 2012. Human pre-implantation embryo development. *Development,* 139**,** 829-41.
NOJIMA, T., GOMES, T., GROSSO, A. R. F., KIMURA, H., DYE, M. J., DHIR, S., CARMO-FONSECA, M. & PROUDFOOT, N. J. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell,* 161**,** 526-540.
NORRIS, D. P., BROCKDORFF, N. & RASTAN, S. 1991. Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mamm Genome,* 1**,** 78-83.
O'NEILL, L. P., SPOTSWOOD, H. T., FERNANDO, M. & TURNER, B. M. 2008. Differential loss of histone H3 isoforms mono-, di- and tri-methylated at lysine 4 during X-inactivation in female embryonic stem cells. *Biol Chem,* 389**,** 365-70.
O'NEILL, M. J., INGRAM, R. S., VRANA, P. B. & TILGHMAN, S. M. 2000. Allelic expression of IGF2 in marsupials and birds. *Dev Genes Evol,* 210**,** 18-20.
OHHATA, T., HOKI, Y., SASAKI, H. & SADO, T. 2008. Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. *Development,* 135**,** 227-35.
OHINATA, Y., PAYER, B., O'CARROLL, D., ANCELIN, K., ONO, Y., SANO, M., BARTON, S. C., OBUKHANYCH, T., NUSSENZWEIG, M., TARAKHOVSKY, A., SAITOU, M. & SURANI, M. A. 2005. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature,* 436**,** 207-13.
OHNO, S. 1967. *Sex chromosomes and sex-linked genes,* Berlin, Spring-Verlag.
OKADA, Y., YAMAGATA, K., HONG, K., WAKAYAMA, T. & ZHANG, Y. 2010. A role for the elongator complex in zygotic paternal genome demethylation. *Nature,* 463**,** 554-8.
OKAE, H., CHIBA, H., HIURA, H., HAMADA, H., SATO, A., UTSUNOMIYA, T., KIKUCHI, H., YOSHIDA, H., TANAKA, A., SUYAMA, M. & ARIMA, T. 2014. Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet,* 10**,** e1004868.
OKAMOTO, I., OTTE, A. P., ALLIS, C. D., REINBERG, D. & HEARD, E. 2004. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science,* 303**,** 644-9.

OKAMOTO, I., PATRAT, C., THEPOT, D., PEYNOT, N., FAUQUE, P., DANIEL, N., DIABANGOUAYA, P., WOLF, J. P., RENARD, J. P., DURANTHON, V. & HEARD, E. 2011. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature,* 472**,** 370-4.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

OKANO, M., XIE, S. & LI, E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet,* 19**,** 219-20.

OOI, S. K. & BESTOR, T. H. 2008. The colorful history of active DNA demethylation. *Cell,* 133**,** 1145-8.

ORTEGA-RECALDE, O., DAY, R. C., GEMMELL, N. J. & HORE, T. A. 2019. Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. *Nat Commun,* 10**,** 3053.

ORTEGA-RECALDE, O. & HORE, T. A. 2019. DNA methylation in the vertebrate germline: balancing memory and erasure. *Essays Biochem,* 63**,** 649-661.

OSWALD, J., ENGEMANN, S., LANE, N., MAYER, W., OLEK, A., FUNDELE, R., DEAN, W., REIK, W. & WALTER, J. 2000. Active demethylation of the paternal genome in the mouse zygote. *Current Biology,* 10**,** 475-478.

OTANI, J., NANKUMO, T., ARITA, K., INAMOTO, S., ARIYOSHI, M. & SHIRAKAWA, M. 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep,* 10**,** 1235-41.

OTTEN, A. B., THEUNISSEN, T. E., DERHAAG, J. G., LAMBRICHS, E. H., BOESTEN, I. B., WINANDY, M., VAN MONTFOORT, A. P., TARBASHEVICH, K., RAZ, E., GERARDS, M., VANOEVELEN, J. M., VAN DEN BOSCH, B. J., MULLER, M. & SMEETS, H. J. 2016. Differences in Strength and Timing of the mtDNA Bottleneck between Zebrafish Germline and Non-germline Cells. *Cell Rep,* 16**,** 622-30.

PAGE, J., BERRIOS, S., RUFAS, J. S., PARRA, M. T., SUJA, J. A., HEYTING, C. & FERNANDEZ-DONOSO, R. 2003. The pairing of X and Y chromosomes during meiotic prophase in the marsupial species Thylamys elegans is maintained by a dense plate developed from their axial elements. *J Cell Sci,* 116**,** 551-60.

PAGE, J., VIERA, A., PARRA, M. T., DE LA FUENTE, R., SUJA, J. A., PRIETO, I., BARBERO, J. L., RUFAS, J. S., BERRIOS, S. & FERNANDEZ-DONOSO, R. 2006. Involvement of synaptonemal complex proteins in sex chromosome segregation during marsupial male meiosis. *PLoS Genet,* 2**,** e136.

PARK, J. S., LEE, D., CHO, S., SHIN, S. T. & KANG, Y. K. 2010. Active loss of DNA methylation in two-cell stage goat embryos. *Int J Dev Biol,* 54**,** 1323-8.

PAYER, B., SAITOU, M., BARTON, S. C., THRESHER, R., DIXON, J. P., ZAHN, D., COLLEDGE, W. H., CARLTON, M. B., NAKANO, T. & SURANI, M. A. 2003. Stella is a maternal effect gene required for normal early development in mice. *Curr Biol,* 13**,** 2110-7.

PEASTON, A. E., EVSIKOV, A. V., GRABER, J. H., DE VRIES, W. N., HOLBROOK, A. E., SOLTER, D. & KNOWLES, B. B. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell,* 7**,** 597-606.

PEAT, J. R., DEAN, W., CLARK, S. J., KRUEGER, F., SMALLWOOD, S. A., FICZ, G., KIM, J. K., MARIONI, J. C., HORE, T. A. & REIK, W. 2014. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. *Cell Rep,* 9**,** 1990-2000.

PENNY, G. D., KAY, G. F., SHEARDOWN, S. A., RASTAN, S. & BROCKDORFF, N. 1996. Requirement for Xist in X chromosome inactivation. *Nature,* 379**,** 131-7.

PERTEA, M., PERTEA, G. M., ANTONESCU, C. M., CHANG, T. C., MENDELL, J. T. & SALZBERG, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol,* 33**,** 290-5.

PESSIA, E., MAKINO, T., BAILLY-BECHET, M., MCLYSAGHT, A. & MARAIS, G. A. 2012. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc Natl Acad Sci U S A,* 109**,** 5346-51.

PETERS, A. H., KUBICEK, S., MECHTLER, K., O'SULLIVAN, R. J., DERIJCK, A. A., PEREZ-BURGOS, L., KOHLMAIER, A., OPRAVIL, S., TACHIBANA, M., SHINKAI, Y., MARTENS, J. H. & JENUWEIN, T. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell,* 12**,** 1577-89.

PETERS, A. H., MERMOUD, J. E., O'CARROLL, D., PAGANI, M., SCHWEIZER, D., BROCKDORFF, N. & JENUWEIN, T. 2002. Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet,* 30**,** 77-80.

PFEIFER, G. P., STEIGERWALD, S. D., HANSEN, R. S., GARTLER, S. M. & RIGGS, A. D. 1990. Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc Natl Acad Sci U S A,* 87**,** 8252-6.

PLATH, K., FANG, J., MLYNARCZYK-EVANS, S. K., CAO, R., WORRINGER, K. A., WANG, H., DE LA CRUZ, C. C., OTTE, A. P., PANNING, B. & ZHANG, Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science,* 300**,** 131-5.

POLANSKI, Z., MOTOSUGI, N., TSURUMI, C., HIIRAGI, T. & HOFFMANN, S. 2008. Hypomethylation of paternal DNA in the late mouse zygote is not essential for development. *Int J Dev Biol,* 52**,** 295-8.

POLLACK, Y., STEIN, R., RAZIN, A. & CEDAR, H. 1980. Methylation of foreign DNA sequences in eukaryotic cells. *Proc Natl Acad Sci U S A,* 77**,** 6463-7.

POPP, C., DEAN, W., FENG, S., COKUS, S. J., ANDREWS, S., PELLEGRINI, M., JACOBSEN, S. E. & REIK, W. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature,* 463**,** 1101-5.

POTOK, M. E., NIX, D. A., PARNELL, T. J. & CAIRNS, B. R. 2013. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell,* 153**,** 759-72.

PRANTERA, G. & FERRARO, M. 1990. Analysis of methylation and distribution of CpG sequences on human active and inactive X chromosomes by in situ nick translation. *Chromosoma,* 99**,** 18-23.

PROUDHON, C., DUFFIE, R., AJJAN, S., COWLEY, M., IRANZO, J., CARBAJOSA, G., SAADEH, H., HOLLAND, M. L., OAKEY, R. J., RAKYAN, V. K., SCHULZ, R. & BOURC'HIS, D. 2012. Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Mol Cell,* 47**,** 909-20.

QUADRANA, L. & COLOT, V. 2016. Plant Transgenerational Epigenetics. *Annu Rev Genet,* 50**,** 467-491.

QUENNEVILLE, S., VERDE, G., CORSINOTTI, A., KAPOPOULOU, A., JAKOBSSON, J., OFFNER, S., BAGLIVO, I., PEDONE, P. V., GRIMALDI, G., RICCIO, A. & TRONO, D. 2011. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell,* 44**,** 361-72.

RAO, S. S., HUNTLEY, M. H., DURAND, N. C., STAMENOVA, E. K., BOCHKOV, I. D., ROBINSON, J. T., SANBORN, A. L., MACHOL, I., OMER, A. D., LANDER, E. S. & AIDEN, E. L. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell,* 159**,** 1665-80.

RAO, S. S. P., HUANG, S. C., GLENN ST HILAIRE, B., ENGREITZ, J. M., PEREZ, E. M., KIEFFER-KWON, K. R., SANBORN, A. L., JOHNSTONE, S. E., BASCOM, G. D., BOCHKOV, I. D., HUANG, X., SHAMIM, M. S., SHIN, J., TURNER, D., YE, Z., OMER, A. D., ROBINSON, J. T., SCHLICK, T., BERNSTEIN, B. E., CASELLAS, R.,

LANDER, E. S. & AIDEN, E. L. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell,* 171**,** 305-320 e24.

REHAN, S. M., GLASTAD, K. M., LAWSON, S. P. & HUNT, B. G. 2016. The Genome and Methylome of a Subsocial Small Carpenter Bee, Ceratina calcarata. *Genome Biol Evol,* 8**,** 1401-10.

REIK, W., DEAN, W. & WALTER, J. 2001. Epigenetic reprogramming in mammalian development. *Science,* 293**,** 1089-93.

REIK, W. & WALTER, J. 2001a. Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. *Nat Genet,* 27**,** 255-6.

REIK, W. & WALTER, J. 2001b. Genomic imprinting: parental influence on the genome. *Nat Rev Genet,* 2**,** 21-32.

REIS SILVA, A. R., ADENOT, P., DANIEL, N., ARCHILLA, C., PEYNOT, N., LUCCI, C. M., BEAUJEAN, N. & DURANTHON, V. 2011. Dynamics of DNA methylation levels in maternal and paternal rabbit genomes after fertilization. *Epigenetics,* 6**,** 987-93.

REN, R., HORTON, J. R., ZHANG, X., BLUMENTHAL, R. M. & CHENG, X. 2018. Detecting and interpreting DNA methylation marks. *Curr Opin Struct Biol,* 53**,** 88-99.

RENS, W., GRUTZNER, F., O'BRIEN P, C., FAIRCLOUGH, H., GRAVES, J. A. & FERGUSON-SMITH, M. A. 2004. Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. *Proc Natl Acad Sci U S A,* 101**,** 16257-61.

RENS, W., WALLDUCK, M. S., LOVELL, F. L., FERGUSON-SMITH, M. A. & FERGUSON-SMITH, A. C. 2010. Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation. *Proc Natl Acad Sci U S A,* 107**,** 17657-62.

RETIEF, J. D., KRAJEWSKI, C., WESTERMAN, M., WINKFEIN, R. J. & DIXON, G. H. 1995. Molecular phylogeny and evolution of marsupial protamine P1 genes. *Proc Biol Sci,* 259**,** 7-14.

RIGGS, A. D. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet,* 14**,** 9-25.

ROBINSON, E. S., SAMOLLOW, P. B., VANDEBERG, J. L. & JOHNSTON, P. G. 1994. X-chromosome replication patterns in adult, newborn and prenatal opossums. *Reprod Fertil Dev,* 6**,** 533-40.

RODRIGUEZ-TERRONES, D. & TORRES-PADILLA, M. E. 2018. Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends Genet,* 34**,** 806-820.

ROSS, M. T., GRAFHAM, D. V., COFFEY, A. J., SCHERER, S., MCLAY, K., MUZNY, D., PLATZER, M., HOWELL, G. R., BURROWS, C., BIRD, C. P., FRANKISH, A., LOVELL, F. L., HOWE, K. L., ASHURST, J. L., FULTON, R. S., SUDBRAK, R., WEN, G., JONES, M. C., HURLES, M. E., ANDREWS, T. D., SCOTT, C. E., SEARLE, S., RAMSER, J., WHITTAKER, A., DEADMAN, R., CARTER, N. P., HUNT, S. E., CHEN, R., CREE, A., GUNARATNE, P., HAVLAK, P., HODGSON, A., METZKER, M. L., RICHARDS, S., SCOTT, G., STEFFEN, D., SODERGREN, E., WHEELER, D. A., WORLEY, K. C., AINSCOUGH, R., AMBROSE, K. D., ANSARI-LARI, M. A., ARADHYA, S., ASHWELL, R. I., BABBAGE, A. K., BAGGULEY, C. L., BALLABIO, A., BANERJEE, R., BARKER, G. E., BARLOW, K. F., BARRETT, I. P., BATES, K. N., BEARE, D. M., BEASLEY, H., BEASLEY, O., BECK, A., BETHEL, G., BLECHSCHMIDT, K., BRADY, N., BRAY-ALLEN, S., BRIDGEMAN, A. M., BROWN, A. J., BROWN, M. J., BONNIN, D., BRUFORD, E. A., BUHAY, C., BURCH, P., BURFORD, D., BURGESS, J., BURRILL, W., BURTON, J., BYE, J. M., CARDER, C., CARREL, L., CHAKO, J., CHAPMAN, J. C., CHAVEZ, D., CHEN, E., CHEN, G., CHEN, Y., CHEN, Z., CHINAULT, C., CICCODICOLA, A., CLARK, S. Y., CLARKE, G., CLEE, C. M., CLEGG, S., CLERC-BLANKENBURG, K., CLIFFORD, K., COBLEY, V., COLE, C. G., CONQUER, J. S., CORBY, N., CONNOR, R. E., DAVID, R., DAVIES, J., DAVIS,

C., DAVIS, J., DELGADO, O., DESHAZO, D., et al. 2005. The DNA sequence of the human X chromosome. *Nature,* 434**,** 325-37.

ROUGIER, N., BOURC'HIS, D., GOMES, D. M., NIVELEAU, A., PLACHOT, M., PALDI, A. & VIEGAS-PEQUIGNOT, E. 1998. Chromosome methylation patterns during mammalian preimplantation development. *Genes & Development,* 12**,** 2108-2113.

RUDENKO, A., DAWLATY, M. M., SEO, J., CHENG, A. W., MENG, J., LE, T., FAULL, K. F., JAENISCH, R. & TSAI, L. H. 2013. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron,* 79**,** 1109-1122.

RUZOV, A., TSENKINA, Y., SERIO, A., DUDNAKOVA, T., FLETCHER, J., BAI, Y., CHEBOTAREVA, T., PELLS, S., HANNOUN, Z., SULLIVAN, G., CHANDRAN, S., HAY, D. C., BRADLEY, M., WILMUT, I. & DE SOUSA, P. 2011. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res,* 21**,** 1332-42.

SADO, T., FENNER, M. H., TAN, S. S., TAM, P., SHIODA, T. & LI, E. 2000. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol,* 225**,** 294-303.

SADO, T., WANG, Z., SASAKI, H. & LI, E. 2001. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development,* 128**,** 1275-86.

SAITO, M. & ISHIKAWA, F. 2002. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem,* 277**,** 35434-9.

SAKAI, Y., SUETAKE, I., SHINOZAKI, F., YAMASHINA, S. & TAJIMA, S. 2004. Co-expression of de novo DNA methyltransferases Dnmt3a2 and Dnmt3L in gonocytes of mouse embryos. *Gene Expr Patterns,* 5**,** 231-7.

SALVAING, J., AGUIRRE-LAVIN, T., BOULESTEIX, C., LEHMANN, G., DEBEY, P. & BEAUJEAN, N. 2012. 5-Methylcytosine and 5-hydroxymethylcytosine spatiotemporal profiles in the mouse zygote. *PLoS One,* 7**,** e38156.

SAMOLLOW, P. B., JOHNSTON, P. G., FORD, A. L. & VANDEBERG, J. L. 1989. X-linked gene expression in metatherian fibroblasts: evidence from the Gpd and Pgk-A loci of the Virginia opossum and the red-necked wallaby. *Biochem Genet,* 27**,** 313-20.

SANFORD, J. P., CLARK, H. J., CHAPMAN, V. M. & ROSSANT, J. 1987. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes & Development,* 1**,** 1039-1046.

SANGRITHI, M. N., ROYO, H., MAHADEVAIAH, S. K., OJARIKRE, O., BHAW, L., SESAY, A., PETERS, A. H., STADLER, M. & TURNER, J. M. 2017. Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. *Dev Cell,* 40**,** 289-301 e3.

SANTOS, F., HENDRICH, B., REIK, W. & DEAN, W. 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol,* 241**,** 172-82.

SANTOS, F., PEAT, J., BURGESS, H., RADA, C., REIK, W. & DEAN, W. 2013. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin,* 6**,** 39.

SANTOS, F., PETERS, A. H., OTTE, A. P., REIK, W. & DEAN, W. 2005. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev Biol,* 280**,** 225-36.

SARDA, S., ZENG, J., HUNT, B. G. & YI, S. V. 2012. The evolution of invertebrate gene body methylation. *Mol Biol Evol,* 29**,** 1907-16.

SASAI, N., NAKAO, M. & DEFOSSEZ, P. A. 2010. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res,* 38**,** 5015-22.

SCHEMPP, W. & MEER, B. 1983. Cytologic evidence for three human X-chromosomal segments escaping inactivation. *Hum Genet,* 63**,** 171-4.

SCHNEIDER, L. K. & RIEKE, W. O. 1967. DNA replication patterns and chromosomal protein synthesis in opossum lymphocytes in vitro. *J Cell Biol,* 33**,** 497-509.

SCHNEIDER, R., BANNISTER, A. J., MYERS, F. A., THORNE, A. W., CRANE-ROBINSON, C. & KOUZARIDES, T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol,* 6**,** 73-7.

SCHULTZ, M. D., HE, Y., WHITAKER, J. W., HARIHARAN, M., MUKAMEL, E. A., LEUNG, D., RAJAGOPAL, N., NERY, J. R., URICH, M. A., CHEN, H., LIN, S., LIN, Y., JUNG, I., SCHMITT, A. D., SELVARAJ, S., REN, B., SEJNOWSKI, T. J., WANG, W. & ECKER, J. R. 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature,* 523**,** 212-6.

SCHULZ, R., WOODFINE, K., MENHENIOTT, T. R., BOURC'HIS, D., BESTOR, T. & OAKEY, R. J. 2008. WAMIDEX: a web atlas of murine genomic imprinting and differential expression. *Epigenetics,* 3**,** 89-96.

SEISENBERGER, S., ANDREWS, S., KRUEGER, F., ARAND, J., WALTER, J., SANTOS, F., POPP, C., THIENPONT, B., DEAN, W. & REIK, W. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell,* 48**,** 849-62.

SEISENBERGER, S., PEAT, J. R., HORE, T. A., SANTOS, F., DEAN, W. & REIK, W. 2013. Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20110330.

SEKI, Y., HAYASHI, K., ITOH, K., MIZUGAKI, M., SAITOU, M. & MATSUI, Y. 2005. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol,* 278**,** 440-58.

SELWOOD, L. 1992. Mechanisms underlying the development of pattern in marsupial embryos. *Curr Top Dev Biol,* 27**,** 175-233.

SELWOOD, L. 2000. Marsupial Egg and Embryo Coats. *Cells Tissues Organs,* 166**,** 208-219.

SELWOOD, L. & JOHNSON, M. H. 2006. Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: evolution and origins. *Bioessays,* 28**,** 128-45.

SELWOOD, L., ROBINSON, E. S., PEDERSEN, R. A. & VANDEBERG, J. L. 1997. Development in vitro of Marsupials: a comparative review of species and a timetable of cleavage and early blastocyst stages of development in Monodelphis domestica. *Int J Dev Biol,* 41**,** 397-410.

SHAPIRO, L. J. 1979. Steroid sulfatase deficiency and X-linked ichthyosis. *Clin Biochem,* 12**,** 205.

SHARIF, J., MUTO, M., TAKEBAYASHI, S., SUETAKE, I., IWAMATSU, A., ENDO, T. A., SHINGA, J., MIZUTANI-KOSEKI, Y., TOYODA, T., OKAMURA, K., TAJIMA, S., MITSUYA, K., OKANO, M. & KOSEKI, H. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature,* 450**,** 908-12.

SHARMAN, G. B. 1971. Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature,* 230**,** 231-2.

SHARP, A. J., STATHAKI, E., MIGLIAVACCA, E., BRAHMACHARY, M., MONTGOMERY, S. B., DUPRE, Y. & ANTONARAKIS, S. E. 2011. DNA methylation profiles of human active and inactive X chromosomes. *Genome Res,* 21**,** 1592-600.

SHEARDOWN, S., NORRIS, D., FISHER, A. & BROCKDORFF, N. 1996. The mouse Smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. *Hum Mol Genet,* 5**,** 1355-60.

SHEN, J. C., RIDEOUT, W. M., 3RD & JONES, P. A. 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res,* 22**,** 972-6.

SHEN, L., INOUE, A., HE, J., LIU, Y., LU, F. & ZHANG, Y. 2014. Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell,* 15**,** 459-471.

SHEVCHENKO, A. I., ZAKHAROVA, I. S., ELISAPHENKO, E. A., KOLESNIKOV, N. N., WHITEHEAD, S., BIRD, C., ROSS, M., WEIDMAN, J. R., JIRTLE, R. L., KARAMYSHEVA, T. V., RUBTSOV, N. B., VANDEBERG, J. L., MAZUROK, N. A., NESTEROVA, T. B., BROCKDORFF, N. & ZAKIAN, S. M. 2007. Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials. *Chromosome Res,* 15**,** 127-36.

SHI, W., DIRIM, F., WOLF, E., ZAKHARTCHENKO, V. & HAAF, T. 2004. Methylation reprogramming and chromosomal aneuploidy in in vivo fertilized and cloned rabbit preimplantation embryos. *Biol Reprod,* 71**,** 340-7.

SHIRANE, K., TOH, H., KOBAYASHI, H., MIURA, F., CHIBA, H., ITO, T., KONO, T. & SASAKI, H. 2013. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet,* 9**,** e1003439.

SIENSKI, G., DONERTAS, D. & BRENNECKE, J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell,* 151**,** 964-80.

SILVA, J., MAK, W., ZVETKOVA, I., APPANAH, R., NESTEROVA, T. B., WEBSTER, Z., PETERS, A. H., JENUWEIN, T., OTTE, A. P. & BROCKDORFF, N. 2003. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell,* 4**,** 481-95.

SINGER-SAM, J., GRANT, M., LEBON, J. M., OKUYAMA, K., CHAPMAN, V., MONK, M. & RIGGS, A. D. 1990. Use of a HpaII-polymerase chain reaction assay to study DNA methylation in the Pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation. *Mol Cell Biol,* 10**,** 4987-9.

SINGH, P., LI, A. X., TRAN, D. A., OATES, N., KANG, E. R., WU, X. & SZABO, P. E. 2013. De novo DNA methylation in the male germ line occurs by default but is excluded at sites of H3K4 methylation. *Cell Rep,* 4**,** 205-19.

SKALETSKY, H., KURODA-KAWAGUCHI, T., MINX, P. J., CORDUM, H. S., HILLIER, L., BROWN, L. G., REPPING, S., PYNTIKOVA, T., ALI, J., BIERI, T., CHINWALLA, A., DELEHAUNTY, A., DELEHAUNTY, K., DU, H., FEWELL, G., FULTON, L., FULTON, R., GRAVES, T., HOU, S. F., LATRIELLE, P., LEONARD, S., MARDIS, E., MAUPIN, R., MCPHERSON, J., MINER, T., NASH, W., NGUYEN, C., OZERSKY, P., PEPIN, K., ROCK, S., ROHLFING, T., SCOTT, K., SCHULTZ, B., STRONG, C., TIN-WOLLAM, A., YANG, S. P., WATERSTON, R. H., WILSON, R. K., ROZEN, S. & PAGE, D. C. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature,* 423**,** 825-37.

SKVORTSOVA, K., TARBASHEVICH, K., STEHLING, M., LISTER, R., IRIMIA, M., RAZ, E. & BOGDANOVIC, O. 2019. Retention of paternal DNA methylome in the developing zebrafish germline. *Nat Commun,* 10**,** 3054.

SMALLWOOD, S. A., TOMIZAWA, S., KRUEGER, F., RUF, N., CARLI, N., SEGONDS-PICHON, A., SATO, S., HATA, K., ANDREWS, S. R. & KELSEY, G. 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet,* 43**,** 811-4.

SMIT, A., HUBLEY, R. & GREEN, P. 2013. *RepeatMasker Open-4.0.* [Online]. Available: www.repeatmasker.org [Accessed 2019].

SMITH, Z. D., CHAN, M. M., HUMM, K. C., KARNIK, R., MEKHOUBAD, S., REGEV, A., EGGAN, K. & MEISSNER, A. 2014. DNA methylation dynamics of the human preimplantation embryo. *Nature,* 511**,** 611-5.

SMITH, Z. D., CHAN, M. M., MIKKELSEN, T. S., GU, H., GNIRKE, A., REGEV, A. & MEISSNER, A. 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature,* 484**,** 339-44.

SMITH, Z. D., SHI, J., GU, H., DONAGHEY, J., CLEMENT, K., CACCHIARELLI, D., GNIRKE, A., MICHOR, F. & MEISSNER, A. 2017. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature,* 549**,** 543-547.

SMITS, G., MUNGALL, A. J., GRIFFITHS-JONES, S., SMITH, P., BEURY, D., MATTHEWS, L., ROGERS, J., PASK, A. J., SHAW, G., VANDEBERG, J. L., MCCARREY, J. R., CONSORTIUM, S., RENFREE, M. B., REIK, W. & DUNHAM, I. 2008. Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. *Nat Genet,* 40**,** 971-6.

SOH, Y. Q., ALFOLDI, J., PYNTIKOVA, T., BROWN, L. G., GRAVES, T., MINX, P. J., FULTON, R. S., KREMITZKI, C., KOUTSEVA, N., MUELLER, J. L., ROZEN, S., HUGHES, J. F., OWENS, E., WOMACK, J. E., MURPHY, W. J., CAO, Q., DE JONG, P., WARREN, W. C., WILSON, R. K., SKALETSKY, H. & PAGE, D. C. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell,* 159**,** 800-13.

SONG, J., TEPLOVA, M., ISHIBE-MURAKAMI, S. & PATEL, D. J. 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science,* 335**,** 709-12.

SPLINTER, E., DE WIT, E., NORA, E. P., KLOUS, P., VAN DE WERKEN, H. J., ZHU, Y., KAAIJ, L. J., VAN IJCKEN, W., GRIBNAU, J., HEARD, E. & DE LAAT, W. 2011. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev,* 25**,** 1371-83.

STADLER, M. B., MURR, R., BURGER, L., IVANEK, R., LIENERT, F., SCHOLER, A., VAN NIMWEGEN, E., WIRBELAUER, C., OAKELEY, E. J., GAIDATZIS, D., TIWARI, V. K. & SCHUBELER, D. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature,* 480**,** 490-5.

STANCHEVA, I., EL-MAARRI, O., WALTER, J., NIVELEAU, A. & MEEHAN, R. R. 2002. DNA methylation at promoter regions regulates the timing of gene activation in Xenopus laevis embryos. *Dev Biol,* 243**,** 155-65.

STANCHEVA, I. & MEEHAN, R. R. 2000. Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev,* 14**,** 313-27.

STEVENS, M. E., MAIDENS, P. M., ROBINSON, E. S., VANDEBERG, J. L., PEDERSEN, R. A. & MONK, M. 1988. DNA Methylation in the Developing Marsupial Embryo. *Development,* 103**,** 719-724.

STOGER, R., KUBICKA, P., LIU, C. G., KAFRI, T., RAZIN, A., CEDAR, H. & BARLOW, D. P. 1993. Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell,* 73**,** 61-71.

STRAHL, B. D. & ALLIS, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-5.

STRINGER, J. M., PASK, A. J., SHAW, G. & RENFREE, M. B. 2014. Post-natal imprinting: evidence from marsupials. *Heredity (Edinb),* 113**,** 145-55.

SUETAKE, I., SHINOZAKI, F., MIYAGAWA, J., TAKESHIMA, H. & TAJIMA, S. 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem,* 279**,** 27816-23.

SURANI, M. A. 2001. Reprogramming of genome function through epigenetic inheritance. *Nature,* 414**,** 122-8.

SURANI, M. A. & BARTON, S. C. 1983. Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. *Science,* 222**,** 1034-6.

SURANI, M. A., BARTON, S. C. & NORRIS, M. L. 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature,* 308**,** 548-50.

SUZUKI, M. M. & BIRD, A. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet,* 9**,** 465-76.

SUZUKI, M. M., KERR, A. R., DE SOUSA, D. & BIRD, A. 2007a. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res,* 17**,** 625-31.

SUZUKI, S., ONO, R., NARITA, T., PASK, A. J., SHAW, G., WANG, C., KOHDA, T., ALSOP, A. E., MARSHALL GRAVES, J. A., KOHARA, Y., ISHINO, F., RENFREE, M. B. & KANEKO-ISHINO, T. 2007b. Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet,* 3**,** e55.

SUZUKI, S., RENFREE, M. B., PASK, A. J., SHAW, G., KOBAYASHI, S., KOHDA, T., KANEKO-ISHINO, T. & ISHINO, F. 2005. Genomic imprinting of IGF2, p57(KIP2) and PEG1/MEST in a marsupial, the tammar wallaby. *Mech Dev,* 122**,** 213-22.

SUZUKI, S., SHAW, G. & RENFREE, M. B. 2013. Postnatal epigenetic reprogramming in the germline of a marsupial, the tammar wallaby. *Epigenetics Chromatin,* 6**,** 14.

SVED, J. & BIRD, A. 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc Natl Acad Sci U S A,* 87**,** 4692-6.

TADA, T., OBATA, Y., TADA, M., GOTO, Y., NAKATSUJI, N., TAN, S., KONO, T. & TAKAGI, N. 2000. Imprint switching for non-random X-chromosome inactivation during mouse oocyte growth. *Development,* 127**,** 3101-5.

TADA, T., TADA, M., HILTON, K., BARTON, S. C., SADO, T., TAKAGI, N. & SURANI, M. A. 1998. Epigenotype switching of imprintable loci in embryonic germ cells. *Dev Genes Evol,* 207**,** 551-61.

TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-5.

TAKAGI, N. 1974. Differentiation of X chromosomes in early female mouse embryos. *Exp Cell Res,* 86**,** 127-35.

TAKAGI, N. & SASAKI, M. 1975. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature,* 256**,** 640-2.

TAKAHASHI, N., COLUCCIO, A., THORBALL, C. W., PLANET, E., SHI, H., OFFNER, S., TURELLI, P., IMBEAULT, M., FERGUSON-SMITH, A. C. & TRONO, D. 2019. ZNF445 is a primary regulator of genomic imprinting. *Genes Dev,* 33**,** 49-54.

TANABE, H., MULLER, S., NEUSSER, M., VON HASE, J., CALCAGNO, E., CREMER, M., SOLOVEI, I., CREMER, C. & CREMER, T. 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A,* 99**,** 4424-9.

TANG, W. W., DIETMANN, S., IRIE, N., LEITCH, H. G., FLOROS, V. I., BRADSHAW, C. R., HACKETT, J. A., CHINNERY, P. F. & SURANI, M. A. 2015. A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell,* 161**,** 1453-67.

TEISSANDIER, A. & BOURC'HIS, D. 2017. Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *EMBO J,* 36**,** 1471-1473.

THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B., GARG, K., JOHN, S., SANDSTROM, R., BATES, D., BOATMAN, L., CANFIELD, T. K., DIEGEL, M., DUNN, D., EBERSOL, A. K., FRUM, T., GISTE, E., JOHNSON, A. K., JOHNSON, E. M., KUTYAVIN, T., LAJOIE, B., LEE, B. K., LEE, K., LONDON, D., LOTAKIS, D., NEPH, S., NERI, F., NGUYEN, E. D., QU, H., REYNOLDS, A. P., ROACH, V., SAFI, A., SANCHEZ, M. E., SANYAL, A., SHAFER, A., SIMON, J. M., SONG, L., VONG, S., WEAVER, M., YAN, Y., ZHANG, Z., ZHANG, Z., LENHARD, B., TEWARI, M., DORSCHNER, M. O., HANSEN, R. S., NAVAS, P. A., STAMATOYANNOPOULOS, G., IYER, V. R., LIEB, J. D., SUNYAEV, S. R., AKEY, J. M., SABO, P. J., KAUL, R., FUREY, T. S., DEKKER, J., CRAWFORD, G. E. & STAMATOYANNOPOULOS, J. A. 2012. The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

TIAN, D., SUN, S. & LEE, J. T. 2010. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell,* 143**,** 390-403.

TODD, C. D., DENIZ, O., TAYLOR, D. & BRANCO, M. R. 2019. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *Elife,* 8.

TOMIZAWA, S., KOBAYASHI, H., WATANABE, T., ANDREWS, S., HATA, K., KELSEY, G. & SASAKI, H. 2011. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development,* 138**,** 811-20.

TOPPINGS, M., CASTRO, C., MILLS, P. H., REINHART, B., SCHATTEN, G., AHRENS, E. T., CHAILLET, J. R. & TRASLER, J. M. 2008. Profound phenotypic variation among mice deficient in the maintenance of genomic imprints. *Hum Reprod,* 23**,** 807-18.

TOYOOKA, Y., TSUNEKAWA, N., TAKAHASHI, Y., MATSUI, Y., SATOH, M. & NOCE, T. 2000. Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech Dev,* 93**,** 139-49.

TRIBIOLI, C., TAMANINI, F., PATROSSO, C., MILANESI, L., VILLA, A., PERGOLIZZI, R., MAESTRINI, E., RIVELLA, S., BIONE, S., MANCINI, M. & ET AL. 1992. Methylation and sequence analysis around EagI sites: identification of 28 new CpG islands in XQ24-XQ28. *Nucleic Acids Res,* 20**,** 727-33.

TSUKADA, Y., AKIYAMA, T. & NAKAYAMA, K. I. 2015. Maternal TET3 is dispensable for embryonic development but is required for neonatal growth. *Sci Rep,* 5**,** 15876.

TUCCI, V., ISLES, A. R., KELSEY, G., FERGUSON-SMITH, A. C. & ERICE IMPRINTING, G. 2019. Genomic Imprinting and Physiological Processes in Mammals. *Cell,* 176**,** 952-965.

TURNER, B. M. 2000. Histone acetylation and an epigenetic code. *Bioessays,* 22**,** 836-45.

TYNDALE-BISCOE, C. H. & RENFREE, M. 1987. *Monographs on Marsupial Biology: Reproductive Physiology of Marsupials,* Cambridge, Cambridge University Press.

UEDA, T., ABE, K., MIURA, A., YUZURIHA, M., ZUBAIR, M., NOGUCHI, M., NIWA, K., KAWASE, Y., KONO, T., MATSUDA, Y., FUJIMOTO, H., SHIBATA, H., HAYASHIZAKI, Y. & SASAKI, H. 2000. The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells,* 5**,** 649-59.

ULLMANN, S. L., SHAW, G., ALCORN, G. T. & RENFREE, M. B. 1997. Migration of primordial germ cells to the developing gonadal ridges in the tammar wallaby Macropus eugenii. *J Reprod Fertil,* 110**,** 135-43.

VAN DER AUWERA, G. A., CARNEIRO, M. O., HARTL, C., POPLIN, R., DEL ANGEL, G., LEVY-MOONSHINE, A., JORDAN, T., SHAKIR, K., ROAZEN, D., THIBAULT, J., BANKS, E., GARIMELLA, K. V., ALTSHULER, D., GABRIEL, S. & DEPRISTO, M. A. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics,* 43**,** 11 10 1-11 10 33.

VASSENA, R., BOUE, S., GONZALEZ-ROCA, E., ARAN, B., AUER, H., VEIGA, A. & IZPISUA BELMONTE, J. C. 2011. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development,* 138**,** 3699-709.

VEENSTRA, G. J. & WOLFFE, A. P. 2001. Constitutive genomic methylation during embryonic development of Xenopus. *Biochim Biophys Acta,* 1521**,** 39-44.

VELASCO, G., HUBE, F., ROLLIN, J., NEUILLET, D., PHILIPPE, C., BOUZINBA-SEGARD, H., GALVANI, A., VIEGAS-PEQUIGNOT, E. & FRANCASTEL, C. 2010. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A,* 107**,** 9281-6.

VESELOVSKA, L., SMALLWOOD, S. A., SAADEH, H., STEWART, K. R., KRUEGER, F., MAUPETIT-MEHOUAS, S., ARNAUD, P., TOMIZAWA, S. I., ANDREWS, S. & KELSEY, G. 2015. Deep sequencing and de novo assembly of the mouse oocyte

transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol,* 16**,** 209.

VEYRUNES, F., WATERS, P. D., MIETHKE, P., RENS, W., MCMILLAN, D., ALSOP, A. E., GRUTZNER, F., DEAKIN, J. E., WHITTINGTON, C. M., SCHATZKAMER, K., KREMITZKI, C. L., GRAVES, T., FERGUSON-SMITH, M. A., WARREN, W. & MARSHALL GRAVES, J. A. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res,* 18**,** 965-73.

VIEGAS-PEQUIGNOT, E., DUTRILLAUX, B. & THOMAS, G. 1988. Inactive X chromosome has the highest concentration of unmethylated Hha I sites. *Proc Natl Acad Sci U S A,* 85**,** 7657-60.

VINCENT, J. J., HUANG, Y., CHEN, P. Y., FENG, S., CALVOPINA, J. H., NEE, K., LEE, S. A., LE, T., YOON, A. J., FAULL, K., FAN, G., RAO, A., JACOBSEN, S. E., PELLEGRINI, M. & CLARK, A. T. 2013. Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell,* 12**,** 470-8.

VLACHOGIANNIS, G., NIEDERHUTH, C. E., TUNA, S., STATHOPOULOU, A., VIIRI, K., DE ROOIJ, D. G., JENNER, R. G., SCHMITZ, R. J. & OOI, S. K. T. 2015. The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis. *Cell Rep,* 11**,** 990.

WALSH, C. P., CHAILLET, J. R. & BESTOR, T. H. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet,* 20**,** 116-7.

WANG, J., XIA, Y., LI, L., GONG, D., YAO, Y., LUO, H., LU, H., YI, N., WU, H., ZHANG, X., TAO, Q. & GAO, F. 2013. Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing. *BMC Genomics,* 14**,** 1-12.

WANG, L., ZHANG, J., DUAN, J., GAO, X., ZHU, W., LU, X., YANG, L., ZHANG, J., LI, G., CI, W., LI, W., ZHOU, Q., ALURU, N., TANG, F., HE, C., HUANG, X. & LIU, J. 2014a. Programming and inheritance of parental DNA methylomes in mammals. *Cell,* 157**,** 979-991.

WANG, M., LIN, F., XING, K. & LIU, L. 2017. Random X-chromosome inactivation dynamics in vivo by single-cell RNA sequencing. *BMC Genomics,* 18**,** 90.

WANG, X. & BHANDARI, R. K. 2019a. DNA methylation dynamics during epigenetic reprogramming of medaka embryo. *Epigenetics,* 14**,** 611-622.

WANG, X. & BHANDARI, R. K. 2019b. The dynamics of DNA methylation during epigenetic reprogramming of primordial germ cells in medaka (Oryzias latipes). *Epigenetics***,** 1-16.

WANG, X., DOUGLAS, K. C., VANDEBERG, J. L., CLARK, A. G. & SAMOLLOW, P. B. 2014b. Chromosome-wide profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the opossum, Monodelphis domestica. *Genome Res,* 24**,** 70-83.

WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., PENG, W., ZHANG, M. Q. & ZHAO, K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet,* 40**,** 897-903.

WARNER, C. M. & VERSTEEGH, L. R. 1974. In vivo and in vitro effect of alpha-amanitin on preimplantation mouse embryo RNA polymerase. *Nature,* 248**,** 678-80.

WATERS, S. A., LIVERNOIS, A. M., PATEL, H., O'MEALLY, D., CRAIG, J. M., MARSHALL GRAVES, J. A., SUTER, C. M. & WATERS, P. D. 2018. Landscape of DNA Methylation on the Marsupial X. *Mol Biol Evol,* 35**,** 431-439.

WEBER, M., DAVIES, J. J., WITTIG, D., OAKELEY, E. J., HAASE, M., LAM, W. L. & SCHUBELER, D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet,* 37**,** 853-62.

WEBSTER, K. E., O'BRYAN, M. K., FLETCHER, S., CREWTHER, P. E., AAPOLA, U., CRAIG, J., HARRISON, D. K., AUNG, H., PHUTIKANIT, N., LYLE, R.,

MEACHEM, S. J., ANTONARAKIS, S. E., DE KRETSER, D. M., HEDGER, M. P., PETERSON, P., CARROLL, B. J. & SCOTT, H. S. 2005. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A,* 102**,** 4068-73.

WEIDMAN, J. R., DOLINOY, D. C., MALONEY, K. A., CHENG, J. F. & JIRTLE, R. L. 2006a. Imprinting of opossum Igf2r in the absence of differential methylation and air. *Epigenetics,* 1**,** 49-54.

WEIDMAN, J. R., MALONEY, K. A. & JIRTLE, R. L. 2006b. Comparative phylogenetic analysis reveals multiple non-imprinted isoforms of opossum Dlk1. *Mamm Genome,* 17**,** 157-67.

WEST, J. D., FRELS, W. I., CHAPMAN, V. M. & PAPAIOANNOU, V. E. 1977. Preferential expression of the maternally derived X chromosome in the mouse yolk sac. *Cell,* 12**,** 873-82.

WHEELAN, S. J., CHURCH, D. M. & OSTELL, J. M. 2001. Spidey: a tool for mRNA-to-genomic alignments. *Genome Res,* 11**,** 1952-7.

WIEHLE, L., RADDATZ, G., MUSCH, T., DAWLATY, M. M., JAENISCH, R., LYKO, F. & BREILING, A. 2016. Tet1 and Tet2 Protect DNA Methylation Canyons against Hypermethylation. *Mol Cell Biol,* 36**,** 452-61.

WIGLER, M., LEVY, D. & PERUCHO, M. 1981. The somatic replication of DNA methylation. *Cell,* 24**,** 33-40.

WOLF, S. F., DINTZIS, S., TONIOLO, D., PERSICO, G., LUNNEN, K. D., AXELMAN, J. & MIGEON, B. R. 1984a. Complete concordance between glucose-6-phosphate dehydrogenase activity and hypomethylation of 3' CpG clusters: implications for X chromosome dosage compensation. *Nucleic Acids Res,* 12**,** 9333-48.

WOLF, S. F., JOLLY, D. J., LUNNEN, K. D., FRIEDMANN, T. & MIGEON, B. R. 1984b. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proc Natl Acad Sci U S A,* 81**,** 2806-10.

WOSSIDLO, M., ARAND, J., SEBASTIANO, V., LEPIKHOV, K., BOIANI, M., REINHARDT, R., SCHOLER, H. & WALTER, J. 2010. Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J,* 29**,** 1877-88.

WOSSIDLO, M., NAKAMURA, T., LEPIKHOV, K., MARQUES, C. J., ZAKHARTCHENKO, V., BOIANI, M., ARAND, J., NAKANO, T., REIK, W. & WALTER, J. 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun,* 2**,** 241.

WOURMS, J. P. & CALLARD, I. P. 1992. A retrospect to the symposium on evolution of vivparity in vertebrates. *American Zoologist***,** 251-255.

WU, J., ELLISON, J., SALIDO, E., YEN, P., MOHANDAS, T. & SHAPIRO, L. J. 1994. Isolation and characterization of XE169, a novel human gene that escapes X-inactivation. *Hum Mol Genet,* 3**,** 153-60.

WU, P., GAO, Y., GUO, W. & ZHU, P. 2019. Using local alignment to enhance single-cell bisulfite sequencing data efficiency. *Bioinformatics,* 35**,** 3273-3278.

XIE, W., SCHULTZ, M. D., LISTER, R., HOU, Z., RAJAGOPAL, N., RAY, P., WHITAKER, J. W., TIAN, S., HAWKINS, R. D., LEUNG, D., YANG, H., WANG, T., LEE, A. Y., SWANSON, S. A., ZHANG, J., ZHU, Y., KIM, A., NERY, J. R., URICH, M. A., KUAN, S., YEN, C. A., KLUGMAN, S., YU, P., SUKNUNTHA, K., PROPSON, N. E., CHEN, H., EDSALL, L. E., WAGNER, U., LI, Y., YE, Z., KULKARNI, A., XUAN, Z., CHUNG, W. Y., CHI, N. C., ANTOSIEWICZ-BOURGET, J. E., SLUKVIN, I., STEWART, R., ZHANG, M. Q., WANG, W., THOMSON, J. A., ECKER, J. R. & REN, B. 2013. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell,* 153**,** 1134-48.

XIONG, Y., CHEN, X., CHEN, Z., WANG, X., SHI, S., WANG, X., ZHANG, J. & HE, X. 2010. RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat Genet,* 42**,** 1043-7.

XU, G. L., BESTOR, T. H., BOURC'HIS, D., HSIEH, C. L., TOMMERUP, N., BUGGE, M., HULTEN, M., QU, X., RUSSO, J. J. & VIEGAS-PEQUIGNOT, E. 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature,* 402**,** 187-91.

YAMAGUCHI, S., HONG, K., LIU, R., SHEN, L., INOUE, A., DIEP, D., ZHANG, K. & ZHANG, Y. 2012. Tet1 controls meiosis by regulating meiotic gene expression. *Nature,* 492**,** 443-7.

YAMAZAKI, Y., MANN, M. R., LEE, S. S., MARH, J., MCCARREY, J. R., YANAGIMACHI, R. & BARTOLOMEI, M. S. 2003. Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. *Proc Natl Acad Sci U S A,* 100**,** 12207-12.

YANG, F., BABAK, T., SHENDURE, J. & DISTECHE, C. M. 2010. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res,* 20**,** 614-22.

YANG, X., HAN, H., DE CARVALHO, D. D., LAY, F. D., JONES, P. A. & LIANG, G. 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell,* 26**,** 577-90.

YASUKOCHI, Y., MARUYAMA, O., MAHAJAN, M. C., PADDEN, C., EUSKIRCHEN, G. M., SCHULZ, V., HIRAKAWA, H., KUHARA, S., PAN, X. H., NEWBURGER, P. E., SNYDER, M. & WEISSMAN, S. M. 2010. X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils. *Proc Natl Acad Sci U S A,* 107**,** 3704-9.

YE, J., COULOURIS, G., ZARETSKAYA, I., CUTCUTACHE, I., ROZEN, S. & MADDEN, T. L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* 13**,** 134.

YEN, P. H., ELLISON, J., SALIDO, E. C., MOHANDAS, T. & SHAPIRO, L. 1992. Isolation of a new gene from the distal short arm of the human X chromosome that escapes X-inactivation. *Hum Mol Genet,* 1**,** 47-52.

YEN, P. H., PATEL, P., CHINAULT, A. C., MOHANDAS, T. & SHAPIRO, L. J. 1984. Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. *Proc Natl Acad Sci U S A,* 81**,** 1759-63.

YIN, Y., MORGUNOVA, E., JOLMA, A., KAASINEN, E., SAHU, B., KHUND-SAYEED, S., DAS, P. K., KIVIOJA, T., DAVE, K., ZHONG, F., NITTA, K. R., TAIPALE, M., POPOV, A., GINNO, P. A., DOMCKE, S., YAN, J., SCHUBELER, D., VINSON, C. & TAIPALE, J. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science,* 356.

YODER, J. A., WALSH, C. P. & BESTOR, T. H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet,* 13**,** 335-40.

YOKOMINE, T., HATA, K., TSUDZUKI, M. & SASAKI, H. 2006. Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenet Genome Res,* 113**,** 75-80.

ZAITSEVA, I., ZAITSEV, S., ALENINA, N., BADER, M. & KRIVOKHARCHENKO, A. 2007. Dynamics of DNA-demethylation in early mouse and rat embryos developed in vivo and in vitro. *Mol Reprod Dev,* 74**,** 1255-61.

ZAMUDIO, N., BARAU, J., TEISSANDIER, A., WALTER, M., BORSOS, M., SERVANT, N. & BOURC'HIS, D. 2015. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev,* 29**,** 1256-70.

ZHANG, R. R., CUI, Q. Y., MURAI, K., LIM, Y. C., SMITH, Z. D., JIN, S., YE, P., ROSA, L., LEE, Y. K., WU, H. P., LIU, W., XU, Z. M., YANG, L., DING, Y. Q., TANG, F., MEISSNER, A., DING, C., SHI, Y. & XU, G. L. 2013a. Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell,* 13**,** 237-45.

ZHANG, W., CHEN, F., CHEN, R., XIE, D., YANG, J., ZHAO, X., GUO, R., ZHANG, Y., SHEN, Y., GOKE, J., LIU, L. & LU, X. 2019. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res,* 47**,** 8485-8501.

ZHANG, Y., CASTILLO-MORALES, A., JIANG, M., ZHU, Y., HU, L., URRUTIA, A. O., KONG, X. & HURST, L. D. 2013b. Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Mol Biol Evol,* 30**,** 2588-601.

ZHANG, Y., XIANG, Y., YIN, Q., DU, Z., PENG, X., WANG, Q., FIDALGO, M., XIA, W., LI, Y., ZHAO, Z. A., ZHANG, W., MA, J., XU, F., WANG, J., LI, L. & XIE, W. 2018. Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nat Genet,* 50**,** 96-105.

ZHANG, Y. L., CHEN, T., JIANG, Y., ZHONG, Z. S., LIU, S. Z., HOU, Y., SCHATTEN, H., CHEN, D. Y. & SUN, Q. Y. 2005. Active demethylation of individual genes in intracytoplasmic sperm injection rabbit embryos. *Mol Reprod Dev,* 72**,** 530-3.

ZHU, P., GUO, H., REN, Y., HOU, Y., DONG, J., LI, R., LIAN, Y., FAN, X., HU, B., GAO, Y., WANG, X., WEI, Y., LIU, P., YAN, J., REN, X., YUAN, P., YUAN, Y., YAN, Z., WEN, L., YAN, L., QIAO, J. & TANG, F. 2018. Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat Genet,* 50**,** 12-19.

ZILBERMAN, D., GEHRING, M., TRAN, R. K., BALLINGER, T. & HENIKOFF, S. 2007. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet,* 39**,** 61-9.

ZINN, A. R., BRESSLER, S. L., BEER-ROMERO, P., ADLER, D. A., CHAPMAN, V. M., PAGE, D. C. & DISTECHE, C. M. 1991. Inactivation of the Rps4 gene on the mouse X chromosome. *Genomics,* 11**,** 1097-101.